



(10) **DE 20 2018 006 448 U1** 2020.11.19

(12)

Gebrauchsmusterschrift

(21) Aktenzeichen: **20 2018 006 448.3**
(22) Anmeldetag: **17.08.2018**
(47) Eintragungstag: **14.10.2020**
(45) Bekanntmachungstag im Patentblatt: **19.11.2020**

(51) Int Cl.: **G06T 3/40 (2006.01)**
G06K 9/62 (2006.01)
G01B 11/14 (2006.01)

(30) Unionspriorität:
62/546,711 **17.08.2017** **US**

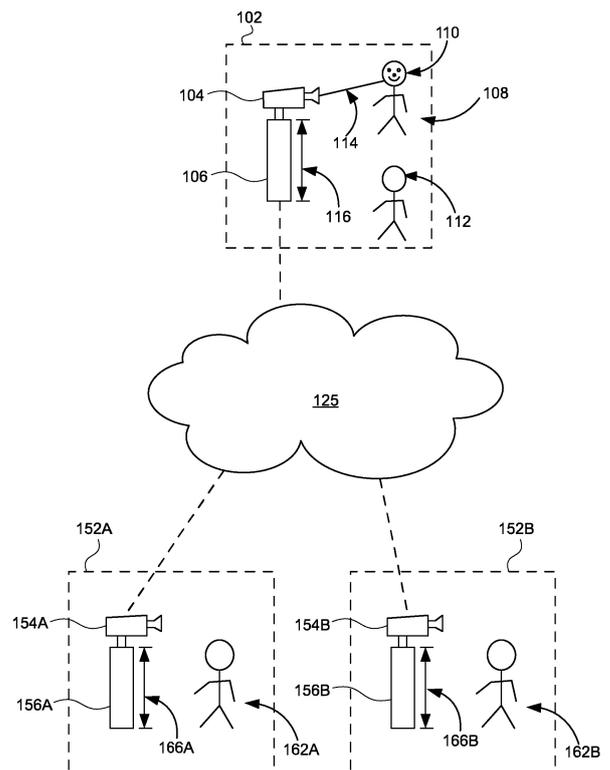
(74) Name und Wohnsitz des Vertreters:
**Betten & Resch Patent- und Rechtsanwälte
PartGmbH, 80333 München, DE**

(73) Name und Wohnsitz des Inhabers:
GOOGLE LLC, Mountain View, Calif., US

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **Skalieren eines Bilds eines Gesichts eines Sprechers basierend auf dem Abstand eines Gesichts und einer Größe einer Anzeige**

(57) Hauptanspruch: Computerprogramm, das Anweisungen umfasst, die dann, wenn sie auf wenigstens einem Prozessor ausgeführt werden, veranlassen, dass ein Computersystem Schritte ausführt, die folgendes umfassen: Bestimmen eines Abstands eines Gesichts einer Person von einer Kamera, die ein Bild des Gesichts der Person aufnahm; Bestimmen einer Größe einer Anzeige in Kommunikation mit der Kamera; und Skalieren des Bilds basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige.



Beschreibung**Figurenliste****TECHNISCHES GEBIET**

[0001] Diese Beschreibung betrifft Videokonferenz.

[0002] Als Gebrauchsmuster geschützt werden und Gegenstand von diesem sind gemäß den Erfordernissen des Gebrauchsmustergesetzes nur Vorrichtungen, wie sie in den beigefügten Ansprüchen definiert sind, aber keine Verfahren. Im Falle, in welchem die Beschreibung auf Verfahren Bezug nimmt, dienen diese Bezugnahmen lediglich dazu, die Vorrichtung oder die Vorrichtungen darzustellen, für welche mit den beigefügten Ansprüchen Schutz gesucht wird.

HINTERGRUND

[0003] Kameras können Bilder eines Raums mit Konferenzteilnehmern aufnehmen, um ein Video zu erzeugen, was eine Videokonferenz ermöglicht. Ein Senden und/oder Präsentieren der unbearbeiteten aufgenommenen Bilder zu und/oder von einer entfernten Anzeige kann in einer Schwierigkeit für Betrachter resultieren, sich auf einen aktuellen Sprecher zu fokussieren, was die „Live“-Gefühl der Videokonferenz reduziert.

ZUSAMMENFASSUNG

[0004] Um die Live-Erfahrung einer Videokonferenz zu verbessern, kann ein Computersystem einen Abstand eines Gesichts eines Sprechenden bzw. Sprechers von einer Kamera während der Videokonferenz bestimmen, eine Größe einer Anzeige bestimmen, auf welcher das Gesicht des Sprechers präsentiert werden wird, und ein aufgenommenes Bild des Gesichts des Sprechers basierend auf dem Abstand und der Größe der Anzeige so skalieren, dass das Bild des Gesichts des Sprechers lebensgroß erscheinen wird. Das angezeigte Bild des Gesichts des Sprechers kann eine selbe Größe haben wie dann, wenn die Anzeige ein Fenster wäre und das Gesicht des Sprechers auf der der Anzeige gegenüberliegenden Seite einen selben Abstand hätte wie das Gesicht des Sprechers von der Kamera. Die Erscheinung des Gesichts des Sprechers in Lebensgröße kann dazu führen, dass sich die Videokonferenz „live“ anfühlt, wobei das Gesicht des Sprechers derart erscheint, dass es der Person (den Personen) nahe ist, die das Gesicht des Sprechers auf der Anzeige anschaut (anschauen).

[0005] Die Details einer oder mehrerer Implementierungen sind in den beigefügten Zeichnungen und der nachstehenden Beschreibung dargelegt. Andere Merkmale werden aus der Beschreibung und den Zeichnungen und aus den Ansprüchen offensichtlich werden.

Fig. 1 ist eine graphische Darstellung eines Systems zum Ermöglichen einer Videokonferenz zwischen Personen in unterschiedlichen Räumen bzw. Zimmern, einschließlich Kameras, Anzeigen und eines Netzwerks, durch das die Kameras und Anzeigen kommunizieren, gemäß einer beispielhaften Implementierung.

Fig. 2 ist ein Blockdiagramm eines Systems zum Skalieren eines Bilds eines Gesichts eines Sprechers gemäß einer beispielhaften Implementierung.

Fig. 3A ist eine graphische Darstellung eines aufgenommenen Bilds, die Bilder von zwei Personen zeigt, gemäß einer beispielhaften Implementierung.

Fig. 3B ist eine graphische Darstellung eines nach einem Skalieren eines Gesichts eines Sprechers präsentierten Bilds gemäß einer beispielhaften Implementierung.

Fig. 4 ist eine graphische Darstellung, die Dimensionen zwischen der Person und der Kamera gemäß einer beispielhaften Implementierung zeigt.

Fig. 5 ist ein Ablaufdiagramm, das ein Verfahren gemäß einer beispielhaften Implementierung zeigt.

Fig. 6 zeigt ein Beispiel einer Computervorrichtung und einer mobilen Computervorrichtung, die verwendet werden können, um die hier beschriebenen Techniken zu implementieren.

DETAILLIERTE BESCHREIBUNG

[0006] Anhand eines Beispiels enthält ein Verfahren ein Bestimmen eines Abstands eines Gesichts einer Person von einer Kamera, die ein Bild des Gesichts der Person aufnahm, ein Bestimmen einer Größe einer Anzeige in Kommunikation mit der Kamera und ein Skalieren des Bilds basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige.

[0007] **Fig. 1** ist eine graphische Darstellung eines Systems zum Ermöglichen einer Videokonferenz zwischen Personen **108**, **112**, **162A**, **162B** in unterschiedlichen Räumen **102**, **152A**, **152B**, einschließlich Kameras **104**, **154A**, **154B**, Anzeigen **106**, **156A**, **156B** und eines Netzwerks **125**, durch das die Kameras **104**, **154A**, **154B** und Anzeigen **106**, **156A**, **156B** kommunizieren, gemäß einer beispielhaften Implementierung. Die Kameras **104**, **154A**, **154B** können Bilder von Teilbereichen der jeweiligen Räume **102**, **152A**, **152B** aufnehmen. Die Teilbereiche der Räume **102**, **152A**, **152B**, für die die Kameras **104**, **154A**, **154B** Bilder aufnahmen, enthalten eine oder mehrere

Personen **108**, **112**, **162A**, **162B**, einschließlich Gesichtern **110** von einer oder mehreren der Personen **108**, **112**, **162A**, **162B**. Die Kameras **104**, **154A**, **154B** können die aufgenommenen Bilder zu den Anzeigen **106**, **156A**, **156B** in den anderen Konferenzräumen **102**, **152A**, **152B** zum Anschauen in den anderen Konferenzräumen **102**, **152A**, **152B** senden. Die Kameras **104**, **154A**, **154B** und Anzeigen **106**, **156A**, **156B** in jedem Raum **102**, **152A**, **152B** können separate Vorrichtungen sein oder können in einzelne Vorrichtungen kombiniert sein.

[0008] Um die aufgenommenen Bilder von Gesichtern, wie beispielsweise dem Gesicht **110** der Person **108** im Raum **102**, auf den die Bilder präsentierenden Anzeigen **156A**, **156B** lebensgroß erscheinen zu lassen, können die Bilder basierend auf einem Abstand **114** des Gesichts **110** von der Kamera **104** und Größen **166A**, **166B** der Anzeigen **156A**, **156B** skaliert werden. Das Skalieren der Bilder kann durch die Kamera **104** und/oder ein die Kamera **104** steuerndes Computersystem, die Anzeigen **156A**, **156B** und/oder die Anzeigen **156A**, **156B** steuernde Computersysteme und/oder durch ein anderes Computersystem (wie beispielsweise einen oder mehrere Server in Kommunikation mit den Kameras **104**, **154A**, **154B** und Anzeigen **106**, **156A**, **156B**), das die Videokonferenz ermöglicht, durchgeführt werden.

[0009] Der einfachen Bezugnahme halber werden Funktionen, Verfahren bzw. Methoden und/oder Techniken hierin unter Bezugnahme auf die Kamera **104** in dem Raum **102** beschrieben werden, die ein Bild der Person **108** und ein Gesicht **110** der Person **108** zur Präsentation auf den Anzeigen **156A**, **156B** in den Räumen **152A**, **152B** aufnimmt. Jedoch können unter Bezugnahme auf die Kamera **104** im Raum **102** in Bezug auf die Personen **108**, **112** beschriebene Funktionen, Verfahren bzw. Methoden und/oder Techniken auch durch die Kameras **154A**, **154B** in Bezug auf die Personen **162A**, **162B** (und/oder irgendwelche anderen Personen in den Räumen **152A**, **152B**) durchgeführt werden und können in Bezug auf die Anzeigen **156A**, **156B** in den Räumen **152A**, **152B** beschriebene Funktionen, Verfahren bzw. Methoden und/oder Techniken auch durch die Anzeige **106** durchgeführt werden. Während zwei Personen **108**, **112** im Raum **102** gezeigt sind und eine Person **162A**, **162B** in jedem der Räume **152A**, **152B** gezeigt ist, kann irgendeine Anzahl von Personen in jedem der Räume **102**, **152A**, **152B** enthalten sein.

[0010] Das das Skalieren durchführende Computersystem kann einen Abstand **114** des Gesichts **110** von der Kamera **104** bestimmen. Das Computersystem kann auch eine Größe **166A**, **166B** der Anzeige **156A**, **156B** bestimmen, die das aufgenommene Bild des Gesichts **110** präsentieren wird. Das Computersystem kann das Bild basierend auf dem bestimm-

ten Abstand und der bestimmten Größe skalieren, um ein lebensgroßes Bild (lebensgroße Bilder) des Gesichts **110** auf der (den) Anzeige(n) **156A**, **156B** zu erzeugen. Das skalierte Bild kann zum Beispiel eine Größe haben, die umgekehrt proportional zu dem bestimmten Abstand des Gesichts **110** von der Kamera **104** ist, so dass das skalierte Bild kleiner sein wird, wenn bestimmt wird, dass das Gesicht **110** weiter von der Kamera **104** weg ist, und größer, wenn bestimmt wird, dass das Gesicht **110** näher zu der Kamera **104** ist. Das skalierte Bild kann zum Beispiel eine Größe haben (wenn sie als ein Anteil der gesamten Größe der Anzeige **156A**, **156B** gemessen wird), die umgekehrt proportional zu der bestimmten Größe der Anzeige(n) **156A**, **156B** ist, so dass das skalierte Bild einen kleineren Anteil einer größeren Anzeige und einen größeren Anteil einer kleineren Anzeige besetzt, so dass die absolute Größe des angezeigten Bilds ungeachtet der Größe der Anzeige(n) **156A**, **156B** dieselbe ist.

[0011] Bei einigen beispielhaften Implementierungen kann das das Skalieren durchführende Computersystem den Abstand **114** des Gesichts **110** der Person **108** von der Kamera **104** basierend auf einem Augenabstand bestimmen. Der Augenabstand kann ein Abstand zwischen Pupillen und/oder Mittelpunkten der Augen des Gesichts **110** der Person **108** sein. Augenabstände neigen dazu, für unterschiedliche Personen nahezu dieselben zu sein, wie beispielsweise etwa fünfundsechzig Millimeter (65 mm). Der Abstand **114** kann bestimmt werden durch Bestimmen des Augenabstands und/oder einer Breite zwischen den Augen auf dem Gesicht **110** der Person **108** und Teilen eines vorbestimmten Abstands oder einer vorbestimmten Breite durch den Augenabstand und/oder der Breite, um ein Abstandsverhältnis und/oder Breitenverhältnis zu bestimmen. Der Abstand **114** des Gesichts **110** von der Kamera **104** kann bestimmt werden durch Multiplizieren des Abstandsverhältnisses und/oder Breitenverhältnisses mit einem vorbestimmten Abstand. Bei einigen Beispielen kann das Computersystem das Abstandsverhältnis und/oder Breitenverhältnis bestimmen durch Teilen des Augenabstands und/oder einer Breite durch den vorbestimmten Abstand oder eine vorbestimmte Breite, und den Abstand **114** bestimmen durch Teilen des vorbestimmten Abstands durch das Abstandsverhältnis und/oder Breitenverhältnis.

[0012] Bei einigen beispielhaften Implementierungen kann das das Skalieren durchführende Computersystem den Abstand basierend auf einer Höhe von einem Mundmittelpunkt zu einem Mittelpunkt zwischen Augen bestimmen. Die Höhe vom Mundmittelpunkt zum Mittelpunkt zwischen Augen neigt dazu, nahezu dieselbe wie der Augenabstand und/oder für unterschiedliche Personen dieselbe zu sein, wie beispielsweise etwa fünfundsechzig Millimeter (65 mm). Das Computersystem kann die Höhe vom

Mundmittelpunkt zum Mittelpunkt zwischen Augen innerhalb des aufgenommenen Bilds messen und eine vorbestimmte Höhe durch die gemessene Höhe teilen, um ein Höhenverhältnis zu bestimmen. Das Computersystem kann den Abstand **114** bestimmen durch Multiplizieren eines vorbestimmten Abstands mit dem Höhenverhältnis. Bei einigen Beispielen kann das Computersystem das Höhenverhältnis bestimmen durch Teilen der gemessenen Höhe durch die vorbestimmte Höhe und den Abstand **114** bestimmen durch Teilen des vorbestimmten Abstands durch das Höhenverhältnis.

[0013] Bei einigen Beispielen kann das Gesicht **110** eher von der Kamera **104** weg, wie beispielsweise zur Seite, nach oben oder nach unten in Bezug auf die Kamera **104**, gerichtet zeigend sein als direkt in Richtung zur Kamera **104**. Das Weggerichtet zeigend sein kann den Augenabstand und/oder die Höhe im aufgenommenen Bild reduzieren. Der Augenabstand und/oder die Höhe im aufgenommenen Bild können oder kann zum Beispiel eine Kosinusfunktion eines Winkels sein, um welchen das Gesicht **110** weg von der Kamera **104** gerichtet zeigend ist. Bei einigen Beispielen kann das Computersystem ein Größeres des Augenabstands und/oder der Breite oder der Höhe vom Mittelpunkt des Mundes zum Mittelpunkt der Augen auswählen. Bei einigen Beispielen kann das Computersystem einen vorbestimmten Wert, wie beispielsweise eine vorbestimmte Breite, einen vorbestimmten Abstand oder eine vorbestimmte Höhe, durch den ausgewählten Wert teilen, um ein Verhältnis zu bestimmen, und das Verhältnis mit einem vorbestimmten Abstand multiplizieren, um den Abstand **114** zu bestimmen. Bei einigen Beispielen kann das Computersystem den ausgewählten Wert durch den vorbestimmten Wert, wie beispielsweise die vorbestimmte Breite, den vorbestimmten Abstand oder die vorbestimmte Höhe, teilen, um das Verhältnis zu bestimmen, und den vorbestimmten Abstand durch das Verhältnis teilen, um den Abstand **114** zu bestimmen. Bei einigen Beispielen kann sich, wenn kein Augenabstand verfügbar ist, wie beispielsweise deshalb, weil die Person zu weit weg von der Kamera **104** ist, das Computersystem auf eine letzte Bestimmung eines Abstands stützen.

[0014] Bei Beispielen, bei welchen das Skalieren durch die Kamera **104** und/ oder ein die Kamera **104** steuerndes Computersystem durchgeführt wird, kann die Kamera **104** und/oder das die Kamera **104** steuernde Computersystem die Größe(n) **166A, 166B** der Anzeige(n) **156A, 156b** basierend auf einer Nachricht oder einem Anzeichen bestimmen, die oder das von der (den) Anzeige(n) **156A, 156B** empfangen ist, wie beispielsweise einen Anzeigengrößenindikator, der in einer Hypertext-Übertragungsprotokoll-(HTTP-)Nachricht enthalten ist, die von der (den) Anzeige(n) **156A, 156B** und/oder einer Computervorrichtung oder einem Computersystem, die oder das mit der

(den) Anzeige(n) **156A, 156B** assoziiert ist oder diese steuert, empfangen ist. Bei Beispielen, bei welchen das Skalieren durch ein Computersystem durchgeführt wird, wie beispielsweise einen oder mehrere Server, die die Videokonferenz ermöglichen, kann das Computersystem die Größe(n) **166A, 166B** der Anzeige(n) **156A, 156B** basierend auf einer Nachricht oder einem Anzeichen bestimmen, die oder das von der (den) Anzeige(n) **156A, 156B** empfangen ist, wie beispielsweise einen Anzeigengrößenindikator, der in einer Hypertext-Übertragungsprotokoll-(HTTP-)Nachricht enthalten ist, die von der (den) Anzeige(n) **156A, 156B** und/oder einer Computervorrichtung oder einem Computersystem, die oder das mit der (den) Anzeige(n) **156A, 156B** assoziiert ist oder diese steuert, empfangen ist. Bei Beispielen, bei welchen das Skalieren durch die Anzeige(n) **156A, 156B** und/oder ein Computersystem, das mit der (den) Anzeige(n) **156A, 156B** assoziiert ist und/oder diese steuert, durchgeführt wird, kann (können) die Anzeige(n) **156A, 156B** die Größe(n) **166A, 166B** basierend auf einem in einem Speicher der Anzeige(n) **156A, 156B** bestimmen.

[0015] Bei einigen Beispielen kann das Computersystem das Gesicht **110** der Person **108** als einen aktiven Sprecher bzw. Sprechenden während der Videokonferenz erkennen. Basierend auf und/oder in Reaktion auf einem Erkennen des Gesichts **110** der Person **108** als den aktiven Sprecher kann das Computersystem das aufgenommene Bild des Gesichts **110** der Person **108** skalieren, das aufgenommene Bild beschneiden, um Teilbereiche zu entfernen, die das Gesicht **110** der Person **108** nicht enthalten, und/oder den Teilbereich des aufgenommenen Bilds, der das Gesicht **110** der Person **108** in einem Zentrum eines erzeugten Bilds und/oder der Anzeige **156A, 156B** enthält, bewegen und/oder präsentieren.

[0016] Um ein Skalieren des Gesichts **110** der Person **108** zu ermöglichen und/oder als eine Vorbedingung oder als Auslöser zum Skalieren des Gesichts **110** der Person **108** kann das Computersystem bestimmen, welche Person **108, 112** der aktive Sprecher ist und/oder die Person **108** und/oder das Gesicht **110** der Person **108** als den aktiven Sprecher in der Videokonferenz erkennen. Bei einigen Beispielen kann das Computersystem bestimmen, dass die Person **108** der aktive Sprecher ist und/oder das Gesicht **110** der Person **108** als den aktiven Sprecher basierend auf einem zeitlichen Unterschied zwischen mehreren Audio-Eingaben, die anzeigen, dass ein Abstand und/oder ein Standort der Person **108** (die als der aktive Sprecher angesehen werden kann) zu einem bestimmten Standort der Person **108** und/oder einem Gesicht **110** der Person **108** passt, erkennen. Bei einigen Beispielen kann der Standort der Person **108** und/oder des Gesichts **110** der Person **108** basierend auf einem bestimmten Abstand der Person **108** und/oder des Gesichts **110** der Person **108** und

einer Ausrichtung der Person **108** und/oder des Gesichts **110** der Person **108** bestimmt werden.

[0017] Bei einigen Beispielen kann das Computersystem die Person **108** und/oder das Gesicht **110** der Person **108** als den aktiven Sprecher basierend auf einem Erkennen einer Lippenbewegung an der Person **108** und/oder dem Gesicht **110** der Person **108** erkennen. Bei einigen Beispielen kann das Computersystem die Person **108** und/oder das Gesicht **110** der Person **108** als den aktiven Sprecher basierend auf einem Erkennen von Gesichtsverformungen im Laufe der Zeit erkennen. Bei einigen Beispielen können Gesichtsverformungen Änderungen an der Erscheinung von Teilen des Gesichts enthalten, einschließlich Augen, Wangen, Kinn und/ oder Stirn. Bei einigen Beispielen kann das Computersystem die Person **108** und/ oder das Gesicht **110** der Person **108** als den aktiven Sprecher basierend auf einem Erkennen von Blicken und/oder Gesichtsverformungen einer anderen Person (von anderen Personen) in Richtung zu der Person **108** und oder dem Gesicht **110** der Person **108** erkennen. Bei einigen Beispielen kann (können) die Person(en), dessen (deren) Blicke derart erkannt werden, dass sie zu der Person **108** und/oder dem Gesicht **110** der Person **108** gerichtet sind, eine oder mehrere Personen **112** im selben Raum **102** wie die Person **108** und/oder das Gesicht **110** der Person **108** sein, die der aktive Sprecher ist und das Bild von deren Gesicht und/oder Augen durch dieselbe Kamera **104** wie die Person **108** und/oder das Gesicht **110** der Person **108** aufgenommen wird. Bei einigen Beispielen kann (können) die Person(en), dessen (deren) Blicke derart erkannt werden, dass sie zu der Person **108** und/oder dem Gesicht **110** der Person **108** gerichtet sind, eine oder mehrere Personen **162A**, **162B** in einem oder mehreren anderen Räumen **152A**, **152B** sein, dessen oder deren Blicke in Richtung zu Bildern der Person **108** und/oder des Gesichts **110** der Person **108** auf der (den) Anzeige(n) **156A**, **156B** durch eine andere Kamera (andere Kameras) **154A**, **154B** als die Kamera **104** aufgenommen werden, die das Bild der Person **108** und oder des Gesichts **110** der Person **108** aufnahm, die der aktive Sprecher ist.

[0018] Bei einigen Beispielen kann das Computersystem die Personen **108**, **112**, **162A**, **162B** basierend auf einer Kombination aus einem, zwei, drei oder vier von einem Erkennen von Gesichtern in aufgenommenen Bildern, einem Erkennen von Körperformen ohne Gesichter in den aufgenommenen Bildern (wie beispielsweise eine Person, die weg von der Kamera **104**, **154A**, **154B** schaut), einem Erkennen von Lippenbewegung und/oder Gesichtsverformungen in den aufgenommenen Bildern und/oder einem Erkennen von Blicken in Richtung zu dem aufgenommenen Bild der Person **108**, **112**, **162A**, **162B** als Personen und/oder Teilnehmer bei der Videokonferenz erkennen, um als eine Person und/oder Teilnehmer bei der Videokonferenz erkannt zu werden.

[0019] Fig. 2 ist ein Blockdiagramm eines Systems **200** zum Skalieren eines Bilds eines Gesichts **110** eines Sprechers gemäß einer beispielhaften Implementierung. Bei einigen Beispielen kann das System **200** in der Kamera **104** sein, und/oder einem Computersystem, das mit der Kamera **200** assoziiert ist und/oder diese steuert. Bei einigen Beispielen kann das System **200** in einem Computersystem, wie beispielsweise einem oder mehreren Servern in Kommunikation mit den Kameras **104**, **154A**, **154B** und/oder Anzeigen **106**, **156A**, **156B**, das die Videokonferenz ermöglicht, enthalten sein. Bei einigen Beispielen kann das System **200** in einer oder mehreren Anzeigen **156A**, **156B** enthalten sein, und/oder einem Computersystem, das mit einer oder mehreren Anzeigen **156A**, **156B** assoziiert ist und diese steuert.

[0020] Das System **200** kann einen Personenerkennung **202** enthalten. Der Personenerkennung **202** kann durch die Kamera(s) **104**, **154A**, **154B** aufgenommene Bilder von Menschen und/oder Personen als Personen und/oder Teilnehmer bei der Videokonferenz erkennen. Der Personenerkennung **202** kann die Bilder als Personen und/oder Teilnehmer bei der Videokonferenz basierend auf Anzeichen bzw. Hinweisen davon, Klassifizierungen als dies und/oder Signalen, die anzeigen, dass die Bilder und/oder Teilbereiche davon Gesichter, Körper, eine Lippenbewegung und/oder Gesichtsverformungen enthalten, und/oder Anzeichen von Blicken in Richtung zu der Person und/oder dem Teilnehmer erkennen.

[0021] Der Personenerkennung **202** kann einen Gesichtserkennung **204** enthalten. Der Gesichtserkennung **204** kann aufgenommene Bilder und/oder Teilbereiche von aufgenommenen Bildern, als Gesichter und/oder menschliche Gesichter erkennen. Der Gesichtserkennung **204** kann Gesichter und/oder menschliche Gesichter basierend auf Erkennungsformen oder Umrissen von Gesichtern und/oder Merkmalen von Gesichtern, wie beispielsweise einem oder mehreren von Augenbraue(n), Auge(n), Nase, Lippe(n) und/oder Kinn, erkennen.

[0022] Der Personenerkennung **202** kann einen Körpererkennung **206** enthalten. Der Körpererkennung **206** kann aufgenommene Bilder und/oder Teilbereiche von aufgenommenen Bildern als menschliche Körper erkennen. Der Körpererkennung **206** kann Körper basierend auf Formen, wie beispielsweise der Form eines Kopfes, eines Halses bzw. Nackens, von Schultern und/oder Brust, die in aufgenommenen Bildern enthalten sind, erkennen.

[0023] Der Personenerkennung **202** kann einen Lippenbewegungserkennung **208** enthalten. Der Lippenbewegungserkennung **208** kann bei aufgenommenen

ensbereich sein können, der Personenerkennung **202** bestimmen, ob eine Person ein Objekt ist, basierend auf einer niedrigeren Anzahl, wie beispielsweise eins oder zwei, von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210**, was mit einem höheren Vertrauensbereich bestimmt, dass das Objekt eine Person ist, wohingegen dann, wenn die vorbestimmte Anzahl, wie beispielsweise eins oder zwei, von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210**, bestimmt, dass eine höhere Anzahl von Objekten Kandidatenpersonen sind, durch Bestimmen, dass die Objekte Personen mit niedrigerem Vertrauensbereich sein können, der Personenerkennung **202** bestimmen kann, ob ein Objekt eine Person ist, basierend auf einer höheren Anzahl, wie beispielsweise drei oder vier, von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210**, was mit einem höheren Vertrauensbereich bestimmt, dass das Objekt eine Person ist.

[0027] Zum Beispiel kann dann, wenn einer oder zwei von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210** bestimmt oder bestimmen, dass eine niedrigere Anzahl, wie beispielsweise nur ein oder zwei Objekte im Raum **102**, Personen mit dem niedrigeren Vertrauensbereich, wie beispielsweise fünfzig Prozent (50%), sind, der Personenerkennung **202** bestimmen, ob das eine oder die zwei Objekte Personen sind, basierend auf der niedrigeren Anzahl, wie beispielsweise eins oder zwei, von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210**, was bestimmt, dass die Objekte Personen mit größerem Vertrauensbereich sind, wie beispielsweise achtzig Prozent (80%). Wenn einer oder zwei von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210** bestimmt oder bestimmen, dass eine größere Anzahl, wie beispielsweise drei oder mehr, von Objekten im Raum **102** Personen mit niedrigerem Vertrauensbereich, wie beispielsweise fünfzig Prozent (50%) sind, dann kann der Personenerkennung **202** bestimmen, ob das eine oder die zwei Objekte Personen sind, basierend auf der höheren Anzahl, wie beispielsweise drei oder vier, von dem Gesichtserkennung **204**, dem Körpererkennung **206**, dem Lippenbewegungserkennung **208** und/oder Gesichtsverformungserkennung und/oder dem Blickbestimmer **210**, was bestimmt, dass die Objekte Per-

sonen mit größerem Vertrauensbereich sind, wie beispielsweise achtzig Prozent (80%).

[0028] Die in den durch die Kamera **104** aufgenommenen Bildern enthaltenen Objekte können Kandidaten sein, um als ein aktiver Sprecher angesehen zu werden. Das System **200** kann das Skalieren an dem Gesicht **110** der Person **108** durchführen, die das System **200** als den aktiven Sprecher bestimmt, und/oder andere Personen **112** in demselben Raum **102** wie der aktive Sprecher.

[0029] Das System kann einer Sprechererkennung **212** enthalten. Der Sprechererkennung **212** kann eine durch den Personenerkennung **202** als einen aktiven Sprecher erkannte Person erkennen. Der Sprechererkennung **212** kann einen aktiven Sprecher durch Bestimmen erkennen, ob jeder der durch den Personenerkennung **202** erkannten Personen ein aktiver Sprecher ist, basierend auf Audiolokalisierung bzw. -standortbestimmung, Lippenbewegung und/oder Gesichtsverformung der erkannten Person, und/oder Blicherfassung von anderen erkannten Personen.

[0030] Der Sprechererkennung **202** kann eine Audiolokalisierungsmaschine **214** enthalten. Die Audiolokalisierungsmaschine **214** kann eine Audiolokalisierung basierend auf erkannter Sprache durchführen, die durch mehrere, wie beispielsweise wenigstens zwei, Mikrofone oder andere Audio-Eingabevorrichtungen in einer Eingabeanordnung, wie beispielsweise einer linearen Mikrofonanordnung, empfangen wird. Die Audiolokalisierungsmaschine **214** kann mögliche Standorte, die Kombinationen von Abstand und/oder Richtung enthalten können, von einem aktiven Sprecher bestimmen, und/oder mehr als einen aktiven Sprecher, wobei die möglichen Standorte einen Bogen bilden, der sich weg von dem Mikrofonen erstreckt, basierend auf Zeitdifferenzen von da an, wenn die Mikrofone dieselbe Sprache empfangen, und einer bekannten Klanggeschwindigkeit. Der Sprechererkennung **212** kann die durch die Audiolokalisierungsmaschine **214** bestimmten möglichen Standorte mit einem durch einen Sprecherstandortbestimmer **220** bestimmten Standort vergleichen, um zu bestimmen, ob eine Person der aktive Sprecher ist.

[0031] Der Sprechererkennung **212** kann einen Lippenbewegungserkennung **216** enthalten. Der Lippenbewegungserkennung **216** kann eine Lippenbewegung von Personen **108**, **112**, wie beispielsweise dem Sprecher, im Raum **102** erkennen. Der Lippenbewegungserkennung **216** kann das Gesicht **110** der Person **108** erkennen und/ oder Lippen innerhalb des Gesichts **110** erkennen. Der Lippenbewegungserkennung **216** kann eine Bewegung der erkannten Lippen erkennen. Basierend auf einer Erkennung einer Lippenbewegung innerhalb eines Gesichts **110** der Person **108** kann der Sprechererkennung **212** bestimmen,

eine Fokussierung und/oder eine Zoomstufe der Kamera **104** steuern. Die Kamera **230** kann die Richtung, die Fokussierung und/oder die Zoomstufe basierend auf dem Standort des Sprechers steuern, wie er durch den Sprecherstandortbestimmer **220** bestimmt ist, so dass das Gesicht und/oder die Brust des Sprechers durch die Kamera **104** aufgenommen werden. Bei einigen Beispielen kann die Kamerasteuerung **230** veranlassen, dass die Kamera **104** in eine Richtung zeigt, so dass die Kamera **104** den Sprecher und eine oder mehrere Personen **112** im selben Raum **102** wie der Sprecher aufnimmt, um ein Erzeugen eines Bilds mit mehreren Personen **108**, **112** zu ermöglichen.

[0039] Bei einigen Beispielen kann die Kamerasteuerung **230** aus der Ferne gesteuert bzw. fernbedient werden, wie beispielsweise durch eine Person **108**, **112** in demselben Raum **102** wie die Kamera **104**, die die Kamerasteuerung **230** steuert, und/oder eine Person **162A**, **162B** in demselben Raum **152A**, **152B** wie die Anzeige **156A**, **156B**, die das Bild (die Bilder) präsentiert, das (die) durch die Kamera **104** aufgenommen ist (sind). Die Kamerasteuerung **230** kann gemäß einigen Beispielen durch eine in der Hand gehaltene Steuerung durch einen Berührungsbildschirm auf der Anzeige **106**, **156A**, **156B** oder durch eine separate Berührungsbildschirmanzeige fernbedient werden. Die Person **108**, **112** kann die Richtung, wie beispielsweise Schwenk und/oder Neigung, und/oder eine Zoomstufe der Kamera **104** steuern. Bei einigen Beispielen kann die Kamerasteuerung **230** einer Person Zoomvorschläge präsentieren, die die Kamera **104** fernbedient, wie beispielsweise einen Vorschlag, eine Zoomstufe auszuwählen, die veranlassen wird, dass ein präsentiertes Bild der Person **108** lebensgroß und/oder naturgetreu erscheint, eine Zoomstufe, die ein breites Bild erzeugt, das alle Personen **108**, **112** im Raum **102** zeigen wird, und/oder eine Zoomstufe, die veranlasst, dass das Bild des Gesichts **110** der Person **108** und/oder andere ausgewählte Körperteile, wie beispielsweise die Brust der Person **108** oder eine Porträtbild, die gesamte Anzeige **156A**, **156B** füllt. Die Kamerasteuerung **230** könnte Regeln und/oder Heuristiken enthalten, um zu verhindern, dass eine Person **108**, **112**, die die Kamera **104** fernbedient, die Kamera **104** auf unerwünschte Standorte, wie beispielsweise ungeeignete Körperteile, fokussiert bzw. ausrichtet. Die ungeeigneten Körperteile könnten durch den Personen-erkenner **202** und/oder den Körpererkenner **206** erkannt und/oder ausgeschlossen werden.

[0040] Das System **200** kann einen Anzeigengrößenbestimmer **232** enthalten. Der Anzeigengrößenbestimmer **232** kann die Größe der Anzeige **156A**, **156B** bestimmen, auf welcher ein Bild, das das Gesicht **110** des Sprechers und/oder Gesichter von irgendwelchen anderen Personen **112** anzeigt, angezeigt werden wird. Bei einigen Beispielen kann das

Bild basierend auf der bestimmten Größe der Anzeige **156A**, **156B** skaliert werden. Wenn die Anzeige **156A**, **156B** kleiner ist, dann kann das Bild einen größeren Anteil der Anzeige **156A**, **156B** besetzen, während dann, wenn die Anzeige **156A**, **156B** größer ist, das Bild einen kleineren Anteil der Anzeige **156A**, **156B** besetzen kann, so dass die Anzeige **156A**, **156B** ein lebensgroßes Bild des Gesichts präsentiert. Bei einem Beispiel, bei welchem ein Computersystem, das mit der Kamera **104** assoziiert ist, die das Bild des Gesichts (der Gesichter) **110** aufnimmt, das Skalieren des Bilds durchführt, kann der Anzeigengrößenbestimmer **232** die Größe der Anzeige **156A**, **156B** basierend auf einer Nachricht bestimmen, die von der Anzeige **156A**, **156B** und/oder einem mit der Anzeige **156A**, **156B** assoziierten Computersystem empfangen wird, wie beispielsweise einer Hypertext-Übertragungsprotokoll-(HTTP-)Nachricht und/oder einer Web-Echtzeitkommunikations-(WebRTC-)Nachricht. Bei einem Beispiel, bei welchem ein mit der Anzeige **156A**, **156B** assoziiertes Computersystem das Skalieren des Bilds durchführt, kann der Anzeigengrößenbestimmer **232** die Größe der Anzeige **156A**, **156B** durch Wiedergewinnen bzw. Auslesen eines Indikators der Größe aus einem Speicher bestimmen, der in der Anzeige **156A**, **156B** und/oder dem mit der Anzeige **156A**, **156B** assoziierten Computersystem gespeichert ist.

[0041] Das Computersystem **200** kann einen Bildgenerator **234** enthalten. Der Bildgenerator **234** kann ein Bild zur Präsentation auf der Anzeige **156A**, **156B** erzeugen. Das erzeugte Bild kann ein Gesicht (Gesichter) **110** von einer oder mehreren Personen **108**, **112**, **162A**, **162B** in den Räumen **102**, **152A**, **152B** enthalten, die an einer Videokonferenz teilnehmen. Der Bildgenerator **234** kann die Bilder skalieren, um die Bilder lebensgroß erscheinen zu lassen, die Bilder so beschneiden, dass die Gesichter **110** auf die Anzeige **106**, **156A**, **156B** passen, und/oder so, dass mehr als ein Gesicht in Lebensgröße auf der Anzeige **106**, **156A**, **156B** angezeigt werden kann, und/oder ein kleineres Bild des Bilds enthalten kann, das durch die Kamera **104**, **154A**, **154B** aufgenommen ist, die in demselben Raum **102**, **152A**, **152B** wie die Anzeige **106**, **156A**, **156B** ist.

[0042] Der Bildgenerator **234** kann einen Skalierer **236** enthalten. Der Skalierer **236** kann Bilder skalieren, und/oder abgeschnittene Teilbereiche von Bildern zur Aufnahme bzw. Einbindung in das Bild, das zur Präsentation auf der Anzeige **156A**, **156B** erzeugt ist. Der Skalierer **236** kann die Bilder basierend auf dem bestimmten Abstand der Person **108**, **112** und/oder des Gesichts **110** und/oder basierend auf der bestimmten Größe der Anzeige **156A**, **156B** skalieren, so dass die Bilder der Gesichter **110** und/oder von Brust lebensgroß erscheinen, als ob die Personen **162A**, **162B**, die auf die durch die Anzeigen **156A**, **156B** präsentierten Bilder schauen, auf

die Person(en) **108, 112** durch ein Fenster schauen und die Person(en) **108, 112** denselben Abstand auf einer gegenüberliegenden Seite der Anzeige **156A, 156B** von der (den) Person(en) **162A, 162B** haben wie die Person(en) **108, 112** von der Anzeige **106**. Bei einigen Beispielen kann der Skalierer **236** Änderungen an der Skalierung verzögern, so dass das Bild nicht ruckartig erscheint, was auftreten könnte, wenn der Skalierer **236** veranlasste, dass das Bild heran- und herauszoomt. Bei einigen Beispielen kann der Skalierer **236** das Bild skalieren, um sicherzustellen, dass „Auffüllungen“ oder Bilder ohne Personen und/oder Teilbereiche des aufgenommenen Bilds, die der Personenerkennung **202** nicht derart erkannt, dass sie irgendeinen Teilbereich einer Person enthalten, auf der linken, rechten und/oder obersten Seite der Anzeige **156A, 156B** existieren. Die Auffüllungen oder Bilder ohne Personen können in einer Anzahl von Pixeln, Inches oder Bruchteilen davon, Zentimetern, Millimetern oder anderen quantifizierbaren Abständen spezifiziert werden.

[0043] Der Bildgenerator **234** kann einen Cropper bzw. Formatwandler **238** enthalten. Der Formatwandler **238** kann das aufgenommene Bild zuschneiden, was Teilbereiche des aufgenommenen Bilds eliminiert und/oder reduziert, wie beispielsweise irgendeine Kombination aus einem linken Teilbereich des aufgenommenen Bilds, einem rechten Teilbereich des aufgenommenen Bilds, einem obersten Teilbereich des aufgenommenen Bilds und/oder einem untersten Teilbereich des aufgenommenen Bilds. Ein Zuschneiden des aufgenommenen Bilds kann ermöglichen, dass der Bildgenerator das aufgenommene Bild durch Erhöhen der Größe des Bilds skaliert, und/oder aufgenommene Bilder von mehr als einem Gesicht **110** und/oder einer Person **108, 112** im erzeugten Bild enthält.

[0044] Das System **200** kann wenigstens einen Prozessor **240** enthalten. Der wenigstens eine Prozessor **240** kann einen oder mehrere Mikroprozessoren und/oder Steuerungen enthalten, die Anweisungen ausführen. Der wenigstens eine Prozessor **240** kann Anweisungen ausführen, um zu veranlassen, dass das System **200** irgendeine Kombination von Funktionen, Verfahren und/oder Techniken durchführt, die hierin in Bezug auf das System **200**, die Kamera(s) **104, 154A, 154B** und/oder das (die) Computersystem(e), das (die) mit der (den) Kamera(s) **104, 154A, 154B** assoziiert ist (sind) und/oder diese steuert (steuern), und/oder die Anzeige(n) **106, 156A, 156B** und/oder das (die) Computersystem(e), das (die) mit der (den) Anzeige(n) **106, 156A, 156B** assoziiert ist (sind) und/oder diese steuert (steuern), beschrieben sind.

[0045] Das System **200** kann wenigstens eine Speichervorrichtung **242** enthalten. Die wenigstens eine Speichervorrichtung **242** kann ein nichtflüchtiges, computerlesbares Speichermedium enthalten, das

Daten und/oder Anweisungen speichert. Die Daten können Daten enthalten, die erzeugt sind und/oder auf die man sich verlässt oder auf die zugegriffen wird während irgendeiner Kombination von Funktionen, Verfahren und/oder Techniken, die hierin beschrieben sind, als ein Prädikat dafür und/oder als Ergebnis davon. Die Anweisungen können Anweisungen enthalten, die dann, wenn sie durch den wenigstens einen Prozessor **240** ausgeführt werden, veranlassen, dass das System **200** irgendeine Kombination von Funktionen, Verfahren und/oder Techniken durchführt, die hierin in Bezug auf das System **200**, die Kamera(s) **104, 154A, 154B** und/oder das (die) Computersystem(e), das (die) mit der (den) Kamera(s) **104, 154A, 154B** assoziiert ist (sind) und/oder diese steuert (steuern), und/oder die Anzeige(n) **106, 156A, 156B** und/oder das (die) Computersystem(e), das (die) mit der (den) Anzeige(n) **106, 156A, 156B** assoziiert ist (sind) und/oder diese steuert (steuern), beschrieben sind.

[0046] Das System **200** kann wenigstens ein Eingabe-/Ausgabemodul **244** enthalten. Das wenigstens eine Eingabe-/Ausgabemodul **244** kann irgendeine Kombination von Eingabe- und/oder Ausgabevorrichtungen enthalten, wie beispielsweise eine oder mehrere Kameras **104, 154A, 154B** (die Videokameras sein können), ein oder mehrere Mikrofone **408A, 408B** (die in **Fig. 4** gezeigt sind), eine oder mehrere Anzeigen **106, 156A, 156B**, eine oder mehrere Human Interface Devices (HIDs), wie beispielsweise eine Tastatur, eine Computermaus oder einen Berührungsbildschirm und/oder Netzwerk- und/oder Kommunikationsschnittstellen einschließlich verdrahteter Schnittstellen (wie beispielsweise Ethernet (Institute for Electrical and Electronics Engineers (IEEE) 802.3), universeller serieller Bus (USB), Koaxialkabel und/oder High Definition Multiple Input (HDMI)), und/oder drahtloser Schnittstellen (wie beispielsweise Wireless Fidelity (IEEE 802.11)), Bluetooth (IEEE 802.15) und/oder eines zellularen Netzwerkprotokolls, wie beispielsweise Long Term Evolution (LTE) und/oder LTE-Advanced), als nicht beschränkende Beispiele.

[0047] **Fig. 3A** ist eine graphische Darstellung eines aufgenommenen Bilds **300**, das Bilder von zwei Personen **308, 312** enthält, gemäß einer beispielhaften Implementierung. Bei einigen Beispielen kann das aufgenommene Bild **300** ein durch die Kamera **104** aufgenommenes Bild sein, und die Personen **308, 312** entsprechen den Personen **108, 112** im Raum **102**, der in **Fig. 1** gezeigt ist. Ein Zentrum **302** des aufgenommenen Bilds **300** kann ein Referenzpunkt zum Bestimmen von Winkeln und/oder Richtungen von Objekten im aufgenommenen Bild **300** sein.

[0048] Das System **200** kann bei wenigstens einer der Personen **308** im aufgenommenen Bild **300** ein Gesicht **310** erkennen. Innerhalb des Gesichts **310**

kann das System Körperteile, wie beispielsweise Augen **322**, **324** und einen Mund **330**, erkennen.

[0049] Bei einigen Beispielen kann das System **200** einen Abstand **326** zwischen den Augen **322**, **324** erkennen. Das System **200** kann den Abstand zwischen Mittelpunkten der Augen, zwischen Pupillen der Augen (der als Augenabstand angesehen werden kann) und/oder zwischen nächsten Punkten der Augen erkennen und/oder bestimmen.

[0050] Bei einigen Beispielen kann das System **200** einen Mittelpunkt **328** der Augen **322**, **324** und einen Mittelpunkt **332** des Mundes **330** erkennen. Das System **200** kann einen Abstand **334** zwischen dem Mittelpunkt **328** der Augen **322**, **324** und dem Mittelpunkt **332** des Mundes **330** erkennen und/oder bestimmen.

[0051] Der Abstandsbestimmer **222** kann den Abstand des Gesichts **310** von der Kamera **104** basierend auf dem Augenabstand **326** und/oder dem Abstand **334** zwischen dem Mittelpunkt **328** der Augen **322**, **324** und dem Mittelpunkt **332** des Mundes **330** bestimmen. Der Abstandsbestimmer **222** kann die Reduzierung in Bezug auf jeden Abstand **326**, **334**, der durch Wegdrehen des Gesichts **310** von der und/oder nicht direktes Schauen in die Kamera **104** kompensieren durch Bestimmen des Abstands des Gesichts **310** von der Kamera **104** basierend auf dem größeren von dem Augenabstand **326** und dem Abstand **334** zwischen dem Mittelpunkt **328** der Augen **322**, **324** und dem Mittelpunkt **332** des Mundes **330**.

[0052] Fig. 3B ist eine graphische Darstellung eines Bilds **350**, das nach einem Skalieren eines Gesichts **310** einer Person präsentiert wird, gemäß einer beispielhaften Implementierung. Bei diesem Beispiel kann der Skalierer **236** den Teilbereich des aufgenommenen Bilds **300**, der das Gesicht **310** enthielt, basierend auf dem (den) bestimmten Abstand (Abständen) **326**, **334** skaliert haben, so dass das Gesicht **310** im Bild **350** auf der Anzeige **156A**, **156B** lebensgroß erscheint (in Fig. 3B nicht gezeigt). Der Formatwandler bzw. Cropper **238** kann das aufgenommene Bild **300** beschnitten haben, um das Gesicht **310** in einem lebensgroßen Maßstab zu präsentieren. Bei dem in Fig. 3B gezeigten Beispiel enthält eine durch den Bildgenerator **234** erzeugtes Bild **350** ein beschnittenes Bild des Gesichts **310** in einem ersten Teilbereich **352** oder linken Teilbereich des Bilds **350** und ein beschnittenes Bild eines Gesichts einer zweiten Person **312** (die der zweiten Person **112** im Raum **102** entsprechen kann) in einem zweiten Teilbereich **354** oder rechten Teilbereich des Bilds **350**.

[0053] Das durch den Bildgenerator **234** erzeugte Bild **350** kann auch ein Miniaturbild **356** des aufgenommenen Bilds **300** und/oder ein Bild des vollständigen Raums **102** enthalten. Das Miniaturbild **356** kann ein Bild des gesamten aufgenommenen Bilds

300 sein, das einen Bruchteil, wie beispielsweise ein Zehntel, des erzeugten Bilds **350** besetzt. Bei einigen Beispielen kann das System **200** das Miniaturbild **356** als ein separates Bild mit niedrigerer Auflösung gegenüber dem übrigen Bild **350** senden. Ein Senden des Miniaturbilds **356** als ein separates Bild mit niedrigerer Auflösung gegenüber dem übrigen Bild **350** kann die Daten reduzieren, die erforderlich sind, um das Bild **350** und das Miniaturbild **356** zu senden, im Vergleich mit einem Senden des Bilds **350** einschließlich des Miniaturbilds **356** als ein einziges Bild oder als zwei Bilder mit derselben Auflösung wie das aufgenommene Bild **350**.

[0054] Fig. 4 ist eine graphische Darstellung, die Dimensionen zwischen der Person **108** und der Kamera **104** gemäß einer beispielhaften Implementierung zeigt. Die Kamera **104** kann einen weiten Winkel haben, was Bilder von Objekten innerhalb von Grenzen **404A**, **404B** aufnimmt, die kleiner als neunzig Grad (90°) von dem Zentrum **302** des Bilds **300** aus, das durch die Kamera **104** aufgenommen ist. Eine Richtung **406** eines Gesichts **110** einer Person **108** kann vom Zentrum **302** des aufgenommenen Bilds **300** aus als Winkel θ gemessen werden. Die Richtung **406** kann basierend auf einem Standort des Gesichts **110** innerhalb des aufgenommenen Bilds **300** bestimmt werden. Das System **200** kann den Winkel θ basierend auf einem Standort des Gesichts **110** innerhalb des aufgenommenen Bilds **300** bestimmen, und/oder basierend auf einem Versatz bzw. Offset des Gesichts **110** vom Zentrum **302** des aufgenommenen Bilds **300**.

[0055] Die Kamera **104** und/oder eine mit der Kamera **104** assoziierte und/oder diese steuernde Computervorrichtung kann mehrere Mikrofone **408A**, **408B** enthalten. Während in Fig. 4 zwei Mikrofone **408A**, **408B** gezeigt sind, kann die Kamera **104** und/oder die Computervorrichtung irgendeine Anzahl von Mikrofonen **408A**, **408B** enthalten. Die Mikrofone **408A**, **408B** können bei diversen und/oder unterschiedlichen Standorten sein. Die diversen und/oder unterschiedlichen Standorte der Mikrofone **408A**, **408B** können veranlassen, dass Abstände **410A**, **410B** von den jeweiligen Mikrofonen **408A**, **408B** zum Gesicht **110** (wo gesprochene Wörter entstehen werden) unterschiedlich voneinander sind. Die unterschiedlichen Abstände **410A**, **410B** vom Gesicht **110** zu den Mikrofonen **408A**, **408B** können veranlassen, dass die durch die Person **108** gesprochenen Wörter und/oder die Audiosignale zu unterschiedlichen Zeiten bei den Mikrofonen **408A**, **408B** ankommen. Die Audiolokalisierungsmaschine **226** des Sprecherstandortbestimmers **220** kann mögliche Abstände und/oder Standorte des Gesichts **110** basierend auf den Unterschieden bzw. Differenzen bezüglich Zeiten zwischen den Mikrofonen **408A**, **408B** bestimmen, die die durch die Person **108** gesprochenen Wörter und/oder Audiosi-

gnale, die von der Person **108** entstehen, empfangen und/oder verarbeiten.

[0056] Fig. 5 ist ein Ablaufdiagramm, das ein Verfahren **500** gemäß einer beispielhaften Implementierung zeigt. Gemäß einem ersten Beispiel kann das Verfahren **500** ein Bestimmen eines Abstands **114** eines Gesichts **110** einer Person von einer Kamera **104** enthalten, die ein Bild **300** des Gesichts **110** der Person **108** aufnahm (**502**), ein Bestimmen einer Größe **166A**, **166B** einer Anzeige **156A**, **156B** in Kommunikation mit der Kamera **104** (**504**) und ein Skalieren des Bilds **350** basierend auf dem bestimmten Abstand **114** des Gesichts **110** der Person **108** und der bestimmten Größe **166A**, **166B** der Anzeige (**506**).

[0057] Gemäß einem zweiten Beispiel kann das Bestimmen des Abstands des Gesichts der Person (**502**) ein Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person, ein Messen einer Breite zwischen dem linken Auge und dem rechten Auge, ein Teilen einer vorbestimmten Breite durch eine gemessene Breite, um ein Breitenverhältnis zu bestimmen, und ein Multiplizieren des Breitenverhältnisses mit einem vorbestimmten Abstand enthalten.

[0058] Gemäß einem dritten Beispiel kann das Bestimmen des Abstands des Gesichts der Person (**502**) ein Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person, ein Bestimmen eines Mittelpunkts zwischen den Augen von dem linken Auge und dem rechten Auge, ein Erkennen eines Mundes im Bild des Gesichts der Person, ein Bestimmen eines Mundmittelpunkts des Mundes, ein Messen einer Höhe vom Mundmittelpunkt zum Mittelpunkt zwischen Augen, ein Teilen einer vorbestimmten Höhe durch die gemessene Höhe, um ein Höhenverhältnis zu bestimmen, und ein Multiplizieren des Höhenverhältnisses mit einem vorbestimmten Abstand enthalten.

[0059] Gemäß einem vierten Beispiel kann das Bestimmen des Abstands des Gesichts der Person (**502**) ein Messen einer Breite zwischen einem linken Auge und einem rechten Auge im Bild des Gesichts der Person, ein Messen einer Höhe zwischen einem Mittelpunkt zwischen Augen und einem Mundmittelpunkt im Bild des Gesichts der Person, ein Teilen einer größeren der Breite und der Höhe durch eine vorbestimmte Breite, wenn die Breite größer ist, und durch eine vorbestimmte Höhe, wenn die Höhe größer ist, um ein Verhältnis zu erzeugen, und ein Multiplizieren des Verhältnisses mit einem vorbestimmten Abstand enthalten.

[0060] Gemäß einem fünften Beispiel kann bei einem vom ersten bis zum vierten Beispiel das Bestimmen der Größe der Anzeige (**504**) ein Bestimmen der Größe der Anzeige basierend auf einem Anzeigen-

größenindikator enthalten, der in einer von einer mit der Anzeige assoziierten Computervorrichtung empfangenen Hyper-text-Übertragungsprotokoll-(HTTP-) Nachricht enthalten ist.

[0061] Gemäß einem sechsten Beispiel kann bei einem von dem ersten bis zum fünften Beispiel das Skalieren des Bilds (**506**) ein Skalieren des Bilds zu einer Größe auf der Anzeige enthalten, die zu einer Größe passt, dass das Gesicht der Person einem Betrachter bei einem Standort der Anzeige aus dem bestimmten Abstand des Gesichts der Person von der Kamera erscheinen würde.

[0062] Gemäß einem siebten Beispiel kann bei einem von dem ersten bis zum sechsten Beispiel das Verfahren **500** weiterhin ein Erkennen des Gesichts der Person als einen aktiven Sprecher basierend auf einer Zeitdifferenz zwischen Audioeingaben enthalten, die anzeigen, dass ein Standort des aktiven Sprechers mit einem bestimmten Standort des Gesichts der Person übereinstimmt.

[0063] Gemäß einem achten Beispiel kann bei einem von dem ersten bis zum siebten Beispiel der Standort des Gesichts der Person basierend auf dem bestimmten Abstand des Gesichts der Person und einer Richtung des Gesichts der Person von der Kamera bestimmt werden.

[0064] Gemäß einem neunten Beispiel kann bei einem von dem ersten bis zum achten Beispiel das Verfahren **500** weiterhin ein Erkennen des Gesichts der Person als einen aktiven Sprecher basierend auf einem Erkennen einer Lippenbewegung des Gesichts der Person enthalten.

[0065] Gemäß einem zehnten Beispiel kann bei einem von dem ersten bis zum neunten Beispiel das Verfahren **500** weiterhin ein Erkennen des Gesichts der Person als einen aktiven Sprecher basierend auf einem Erkennen von Blicken der Augen von anderen Personen in Richtung zu dem Gesicht der Person enthalten.

[0066] Gemäß einem elften Beispiel kann bei einem von dem ersten bis zum zehnten Beispiel das Verfahren **500** weiterhin ein Senden des skalierten Bilds zur Anzeige enthalten.

[0067] Gemäß einem zwölften Beispiel kann bei einem von dem ersten bis zum elften Beispiel das Verfahren **500** weiterhin ein Empfangen des aufgenommenen Bilds von der Kamera enthalten.

[0068] Gemäß einem dreizehnten Beispiel kann bei einem von dem ersten bis zum zwölften Beispiel das Verfahren **500** weiterhin durch ein Computersystem **200** durchgeführt werden, das die Anzeige steuert.

[0069] Gemäß einem vierzehnten Beispiel kann bei einem von dem ersten bis zum dreizehnten Beispiel das Verfahren **500** weiterhin durch ein Computersystem **200** durchgeführt werden, das die Kamera steuert.

[0070] Fig. 6 zeigt ein Beispiel einer gattungsgemäßen bzw. allgemeinen Computervorrichtung **600** und einer gattungsgemäßen bzw. allgemeinen mobilen Computervorrichtung **650**, die mit den hier beschriebenen Techniken verwendet werden können. Die Computervorrichtung **600** soll verschiedene Formen von digitalen Computern darstellen, wie beispielsweise Laptops, Desktops, Tablets, Workstations, persönliche digitale Assistenten bzw. PDAs, Fernsehgeräte, Server, Bladeserver, Großrechner und andere geeignete Computervorrichtungen. Die Computervorrichtung **650** soll verschiedene Formen von mobilen Vorrichtungen darstellen, wie beispielsweise persönliche digitale Assistenten bzw. PDAs, zellulare Telefone, Smartphones und andere ähnliche Computervorrichtungen. Die hier gezeigten Komponenten, ihre Verbindungen und Beziehungen und ihre Funktionen sind nur beispielhaft gemeint und es ist nicht gemeint, dass sie Implementierungen der in diesem Dokument beschriebenen und/oder beanspruchten Erfindungen beschränken.

[0071] Die Computervorrichtung **600** enthält einen Prozessor **602**, einen Speicher **604**, eine Speichervorrichtung **606**, eine Hochgeschwindigkeitsschnittstelle **608**, die mit dem Speicher **604** verbindet, und Hochgeschwindigkeits-Erweiterungsports **610**, und eine Niedergeschwindigkeits-Schnittstelle **612**, die mit einem Niedergeschwindigkeits-Bus **614** und der Speichervorrichtung **606** verbindet. Der Prozessor **602** kann ein halbleiterbasierter Prozessor sein. Jede der Komponenten **602**, **604**, **606**, **608**, **610** und **612** ist unter Verwendung verschiedener Busse miteinander verbunden und kann auf einer gemeinsamen Hauptplatine oder auf andere Weisen, wie es geeignet ist, angebracht sein. Der Prozessor **602** kann Anweisungen zur Ausführung innerhalb der Computervorrichtung **600** verarbeiten, einschließlich Anweisungen, die im Speicher **604** oder auf der Speichervorrichtung **606** gespeichert sind, um graphische Information für eine GUI auf einer externen Eingabe/Ausgabe-Vorrichtung anzuzeigen, wie beispielsweise einer Anzeige **616**, die mit der Hochgeschwindigkeits-Schnittstelle **608** gekoppelt ist. Bei anderen Implementierungen können mehrere Prozessoren und/oder mehrere Busse verwendet werden, wie es geeignet ist, zusammen mit mehreren Speichern und Typen eines Speichers. Zusätzlich können mehrere Computervorrichtungen **600** verbunden sein, wobei jede Vorrichtung Teilbereiche von den nötigen Operationen zur Verfügung stellt (z.B. als eine Serverbank, eine Gruppe von Bladeservern oder ein Mehrprozessorsystem).

[0072] Der Speicher **604** speichert Information innerhalb der Computervorrichtung **600**. Bei einer Implementierung ist der Speicher **604** eine flüchtige Speichereinheit oder -einheiten. Bei einer weiteren Implementierung ist der Speicher **604** eine nichtflüchtige Speichereinheit oder -einheiten. Der Speicher **604** kann auch eine andere Form eines computerlesbaren Mediums sein, wie beispielsweise eine magnetische oder optische Platte.

[0073] Die Speichervorrichtung **606** kann einen Massenspeicher für die Computervorrichtung **600** zur Verfügung stellen. Bei einer Implementierung kann die Speichervorrichtung **606** ein computerlesbares Medium sein oder ein solches enthalten, wie beispielsweise eine Floppydisk-Vorrichtung, eine Festplattenvorrichtung, eine optische Plattenvorrichtung oder eine Bandvorrichtung, ein Flash-Speicher oder eine ähnliche Festkörperspeichervorrichtung oder eine Anordnung von Vorrichtungen, einschließlich Vorrichtungen in einem Speicherbereichsnetzwerk oder anderen Konfigurationen. Ein Computerprogrammprodukt kann greifbar in einem Informationsträger verkörpert sein. Das Computerprogrammprodukt kann auch Anweisungen enthalten, die dann, wenn sie ausgeführt werden, ein oder mehrere Verfahren durchführen, wie beispielsweise diejenigen, die oben beschrieben sind. Der Informationsträger ist ein computer- oder maschinenlesbares Medium, wie beispielsweise der Speicher **604**, die Speichervorrichtung **606** oder ein Speicher auf dem Prozessor **602**.

[0074] Die Hochgeschwindigkeits-Steuerung **608** managt bandbreitenintensive Operationen für die Computervorrichtung **600**, während die Niedergeschwindigkeits-Steuerung **612** Operationen mit niedrigerer Bandbreitenintensität managt. Eine solche Zuteilung von Funktionen ist nur beispielhaft. Bei einer Implementierung ist die Hochgeschwindigkeits-Steuerung **608** mit dem Speicher **604**, der Anzeige **616** (z.B. über einen Grafikprozessor oder einen Akzelerator) und mit Hochgeschwindigkeits-Erweiterungsports **610** gekoppelt, die verschiedene Erweiterungskarten (nicht gezeigt) aufnehmen können. Bei der Implementierung ist die Niedergeschwindigkeits-Steuerung **612** mit der Speichervorrichtung **616** und einem Niedergeschwindigkeits-Erweiterungsport **614** gekoppelt. Das Niedergeschwindigkeits-Erweiterungsport, das verschiedene Kommunikationsports (z.B. USB, Bluetooth, Ethernet, drahtloses Ethernet) enthalten kann, kann mit einer oder mehreren Eingabe/Ausgabe-Vorrichtungen gekoppelt sein, wie beispielsweise einer Tastatur, einer Zeigevorrichtung, einem Scanner oder einer Netzwerkeinrichtung, wie beispielsweise einem Schalter oder einem Router, z.B. über einen Netzwerkadapter.

[0075] Die Computervorrichtung **600** kann auf eine Anzahl von unterschiedlichen Formen implementiert

sein, wie es in der Figur gezeigt ist. Beispielsweise kann sie als standardmäßiger Server **620** implementiert sein, oder mehrere Male in einer Gruppe von solchen Servern. Sie kann auch als Teil eines Rackserversystems **624** implementiert sein. Zusätzlich kann sie in einem Personalcomputer, wie beispielsweise einem Laptop-Computer **622** implementiert sein. Alternativ können Komponenten von der Computervorrichtung **600** mit anderen Komponenten in einer mobilen Vorrichtung (nicht gezeigt) kombiniert sein, wie beispielsweise der Vorrichtung **650**. Jede von solchen Vorrichtungen kann eine oder mehrere von Computervorrichtungen **600**, **650** enthalten, und ein gesamtes System kann aus mehreren Computervorrichtungen **600**, **650** gebildet sein, die miteinander kommunizieren.

[0076] Die Computervorrichtung **650** enthält einen Prozessor **652**, einen Speicher **664**, eine Eingabe/Ausgabe-Vorrichtung, wie beispielsweise eine Anzeige **654**, eine Kommunikationsschnittstelle **666** und einen Transceiver **668**, unter anderen Komponenten. Die Vorrichtung **650** kann auch mit einer Speichervorrichtung versehen sein, wie beispielsweise einem Mikrolaufwerk oder einer anderen Vorrichtung, um zusätzlichen Speicher zur Verfügung zu stellen. Jede der Komponenten **650**, **652**, **664**, **654**, **666** und **668** ist unter Verwendung verschiedener Busse miteinander verbunden, und einige der Komponenten können auf einer gemeinsamen Hauptplatine oder auf anderen Weisen, wie es geeignet ist, angebracht sein.

[0077] Der Prozessor **652** kann Anweisungen innerhalb der Computervorrichtung **650** ausführen, einschließlich Anweisungen, die im Speicher **664** gespeichert sind, der Prozessor kann als eine Chipgruppe von Chips implementiert sein, die separate und mehrere analoge und digitale Prozessoren enthalten. Der Prozessor kann beispielsweise zur Koordination der anderen Komponenten der Vorrichtung **650** sorgen, wie beispielsweise eine Steuerung von Anwenderschnittstellen, Anwendungen, die durch die Vorrichtung **650** laufen, und eine drahtlose Kommunikation durch die Vorrichtung **650**.

[0078] Der Prozessor **652** kann mit einem Anwender über die Steuerungsschnittstelle **658** und die Anzeigeschnittstelle **656**, die mit einer Anzeige **654** gekoppelt ist, kommunizieren. Die Anzeige **654** kann beispielsweise eine TFT LCD (Dünnschichttransistor-Flüssigkristallanzeige) oder eine OLED-(organische lichtemittierende Dioden-)Anzeige oder eine andere geeignete Anzeigetechnologie sein. Die Anzeigeschnittstelle **656** kann eine geeignete Schaltung zum Antreiben der Anzeige **654** umfassen, um einem Anwender graphische und andere Information zu präsentieren. Die Steuerungsschnittstelle **658** kann Befehle von einem Anwender empfangen und sie zur Weiterleitung zum Prozessor **652** umwandeln. Zusätzlich kann eine externe Schnittstelle **662** in Kommunika-

tion mit dem Prozessor **652** zur Verfügung gestellt sein, um eine Nahfeldkommunikation der Vorrichtung **650** mit anderen Vorrichtungen zu ermöglichen. Die externe Schnittstelle **662** kann beispielsweise für eine verdrahtete Kommunikation bei einigen Implementierungen sorgen, oder für eine drahtlose Kommunikation bei anderen Implementierungen, und mehrere Schnittstellen können auch verwendet werden.

[0079] Der Speicher **664** speichert Information innerhalb der Computervorrichtung **650**. Der Speicher **664** kann als ein oder mehrere eines computerlesbaren Mediums oder von Medien implementiert sein, einer flüchtigen Speichereinheit oder Einheiten, oder einer nichtflüchtigen Speichereinheit oder Einheiten. Ein Erweiterungsspeicher **674** kann auch vorgesehen und mit der Vorrichtung **650** über die Erweiterungsschnittstelle **672** verbunden sein, wie beispielsweise eine SIMM-(Einzelreihiges Speichermodul-)Kartenschnittstelle. Ein solcher Erweiterungsspeicher **674** kann zusätzlichen Speicherplatz für die Vorrichtung **650** zur Verfügung stellen oder kann auch Anwendungen oder andere Information für die Vorrichtung **650** speichern. Spezifisch kann der Erweiterungsspeicher **674** Anweisungen enthalten, um die oben beschriebenen Prozesse auszuführen oder zu ergänzen, und kann auch sichere Information enthalten. Somit kann der Erweiterungsspeicher **674** beispielsweise als ein Sicherheitsmodul für die Vorrichtung **650** vorgesehen sein und kann mit Anweisungen programmiert sein, die eine sichere Verwendung der Vorrichtung **650** erlauben. Zusätzlich können sichere Anwendungen über die SIMM-Karten zusammen mit zusätzlicher Information, wie beispielsweise einem Platzieren von identifizierender Information auf der SIMM-Karte auf eine nicht hackbare Weise, vorgesehen sein.

[0080] Der Speicher kann beispielsweise einen Flash-Speicher und/oder einen NVRAM-Speicher enthalten, wie es nachstehend diskutiert ist. Bei einigen Implementierungen ist ein Computerprogrammprodukt greifbar in einem Informationsträger verkörpert bzw. ausgeführt. Das Computerprogrammprodukt enthält Anweisungen, die dann, wenn sie ausgeführt werden, ein oder mehrere Verfahren durchführen, wie beispielsweise diejenigen, die oben beschrieben sind. Der Informationsträger ist ein computer- oder maschinenlesbares Medium, wie beispielsweise der Speicher **664**, der Erweiterungsspeicher **674** oder der Speicher auf dem Prozessor **652**, die beispielsweise über einen Transceiver **668** oder eine externe Schnittstelle **662** aufgenommen werden können.

[0081] Die Vorrichtung **650** kann drahtlos über die Kommunikationsschnittstelle **666** kommunizieren, die eine digitale Signalverarbeitungsschaltung enthalten kann, wo es nötig ist. Die Kommunikationsschnittstelle **666** kann für Kommunikationen un-

ter verschiedenen Moden oder Protokollen sorgen, wie beispielsweise GSM-Sprachaufrufe, SMS, EMS oder MMS-Nachrichtenübermittlung, CDMA, TDMA, PDC, WCDMA, CDMA2000 oder GPRS, und zwar unter anderem. Eine solche Kommunikation kann beispielsweise über einen Funkfrequenztransceiver **668** erfolgen. Zusätzlich kann eine Kurzstreckenkommunikation auftreten, wie beispielsweise unter Verwendung von Bluetooth, Wi-Fi, oder einem anderen solchen Transceiver (nicht gezeigt). Zusätzlich kann ein GPS-(globales Positioniersystem-)Empfängermodul **670** zusätzliche navigations- und standortbezogene drahtlose Daten zu der Vorrichtung **650** liefern, die als geeignet durch Anwendungen verwendet werden können, die auf der Vorrichtung **650** laufen.

[0082] Die Vorrichtung **650** kann auch hörbar unter Verwendung von Audiocodex **660** kommunizieren, was gesprochene Information von einem Anwender empfangen und sie in nutzbare digitale Information umwandeln kann. Audiocodex **660** kann gleichermaßen hörbaren Klang für einen Anwender erzeugen, wie beispielsweise durch einen Lautsprecher, z.B. in einem Handgerät der Vorrichtung **650**. Ein solcher Klang kann Klang von Sprachtelefonanrufen enthalten, kann aufgezeichneten Klang enthalten (z.B. Sprachnachrichten, Musikdateien, etc.) und kann auch Klang enthalten, der durch Anwendungen erzeugt ist, die auf der Vorrichtung **650** in Betrieb sind.

[0083] Die Computervorrichtung **650** kann auf einer Anzahl von unterschiedlichen Formen implementiert sein, wie es in der Figur gezeigt ist. Beispielsweise kann sie als zelluläres Telefon **680** implementiert sein. Sie kann auch als Teil eines Smartphones **682**, eines persönlichen digitalen Assistenten oder einer anderen ähnlichen mobilen Vorrichtung implementiert sein.

[0084] Verschiedene Implementierungen der hier beschriebenen Systeme und Techniken können in einer digitalen elektronischen Schaltung, einer integrierten Schaltung, speziell entwickelten ASICs (anwendungsspezifischen integrierten Schaltungen), Computer-Hardware, -Firmware, -Software und/oder Kombinationen davon realisiert werden. Diese verschiedenen Implementierungen können eine Implementierung in einem oder mehreren Computerprogrammen enthalten, die auf einem programmierbaren System ausführbar und/oder interpretierbar sind, das wenigstens einen programmierbaren Prozessor enthält, der speziell oder allgemein sein kann, gekoppelt, um Daten und Anweisungen von einem Speichersystem zu empfangen und Daten und Anweisungen zu diesem zu senden, wenigstens eine Eingabevorrichtung und wenigstens eine Ausgabevorrichtung.

[0085] Diese Computerprogramme (die auch als Programme, Software, Softwareanwendungen oder Code bekannt sind) enthalten Maschinenanweisungen für einen programmierbaren Prozessor und können in einer höheren verfahrens- und/oder objektorientierten Programmiersprache und/oder in Assembler/Maschinen-Sprache implementiert sein. Wie sie hierin verwendet sind, beziehen sich die Ausdrücke „maschinenlesbares Medium“ „computerlesbares Medium“ auf ein Computerprogrammprodukt, ein Gerät und/oder eine Vorrichtung (z.B. magnetische Platten, optische Platten, einen Speicher, programmierbare Logikvorrichtungen (PLDs)), das oder die verwendet wird, um Maschinenanweisungen und/oder Daten zu einem programmierbaren Prozessor zu liefern, einschließlich eines maschinenlesbaren Mediums, das Maschinenanweisungen als ein maschinenlesbares Signal empfängt. Der Ausdruck „maschinenlesbares Signal“ bezieht sich auf irgendein Signal, das verwendet wird, um Maschinenanweisungen und/oder Daten zu einem programmierbaren Prozessor zu liefern.

[0086] Um für eine Interaktion mit einem Anwender zu sorgen, können die hier beschriebenen Systeme und Techniken auf einem Computer implementiert sein, der eine Anzeigevorrichtung (z.B. einen CRT-(Kathodenstrahlröhren-) oder LCD-(Flüssigkristallanzeigen-)Monitor) zum Anzeigen von Information zum Anwender und eine Tastatur und eine Zeigevorrichtung (z.B. eine Maus oder einen Trackball), wodurch der Anwender eine Eingabe zum Computer liefern kann, hat. Andere Arten von Vorrichtungen können ebenso gut verwendet werden, um für eine Interaktion mit einem Anwender zu sorgen; beispielsweise kann eine zum Anwender gelieferte Rückkopplung irgendeine Form von sensorischer Rückkopplung sein (z.B. eine visuelle Rückkopplung, eine akustische Rückkopplung oder eine taktile Rückkopplung); und eine Eingabe vom Anwender kann in irgendeiner Form empfangen werden, einschließlich einer akustischen, sprachlichem oder taktilen Eingabe.

[0087] Die hier beschriebenen Systeme und Techniken können in einem Computersystem implementiert sein, das eine Backend-Komponente (z.B. als einen Datenserver) enthält oder das eine Middleware-Komponente (z.B. einen Anwendungsserver) enthält oder das eine Frontend-Komponente (z.B. einen Client-Computer mit einer graphischen Anwenderschnittstelle oder einem Web-Browser, durch welchen ein Anwender mit einer Implementierung der hier beschriebenen Systeme und Techniken interagieren kann) enthält, oder irgendeine Kombination von solchen Backend-, Middleware- oder Frontend-Komponenten. Die Komponenten des Systems können durch irgendeine Form oder ein Medium einer digitalen Datenkommunikation (z.B. ein Kommunikationsnetzwerk) miteinander verbunden sein. Beispiele von Kommunikationsnetzwerken enthalten ein loka-

les Netz („LAN“), ein Weitverkehrsnetz („WAN“) und das Internet.

[0088] Das Computersystem kann Clients und Server enthalten. Ein Client und ein Server sind allgemein entfernt voneinander und interagieren typischerweise über ein Kommunikationsnetzwerk. Die Beziehung von Client und Server entsteht mittels Computerprogrammen, die auf den jeweiligen Computern laufen und eine Client-Server-Beziehung zueinander haben.

[0089] Eine Anzahl von Ausführungsformen ist hierin beschrieben worden. Nichtsdestoweniger wird es verstanden werden, dass verschiedene Modifikationen durchgeführt werden können, ohne vom Sinngehalt und Schutzzumfang der Erfindung abzuweichen.

[0090] Zusätzlich erfordern die in den Figuren gezeigten logischen Abläufe nicht die gezeigte besondere Reihenfolge oder sequentielle Reihenfolge, um erwünschte Ergebnisse zu erreichen. Zusätzlich können andere Schritte vorgesehen sein oder können Schritte von den beschriebenen Abläufen eliminiert sein, und andere Komponenten können zu den beschriebenen Systemen hinzugefügt oder davon entfernt sein. Demgemäß sind andere Ausführungsformen innerhalb des Schutzzumfangs der folgenden Ansprüche.

Bezugszeichenliste

102, 152A, 152B =	Raum;
104, 154A, 154B =	Kamera;
106, 156A, 156B =	Anzeige;
108, 112, 162A, 162B =	Person;
110 =	Gesicht;
114 =	Abstand;
116, 166A, 166B =	Größe;
125 =	Netzwerk;
200 =	System;
202 =	Personenerkennung;
204 =	Gesicht;
206 =	Körper;
208, 216 =	Lippenbewegung;
210, 218 =	Blick;
212 =	Sprechererkennung;
214 =	Audiolokalisierung;
220 =	Sprecherstandortbestimmer;
222 =	Abstandsbestimmer;

224 =	Gesichtsmerkmalsanalysator;
226 =	Audiolokalisierung;
228 =	Richtungsbestimmer;
230 =	Kamerasteuerung;
232 =	Anzeigengrößenbestimmer;
234 =	Bildgenerator;
236 =	Skalierer;
238 =	Formatwandler;
240 =	Prozessor;
242 =	Speicher;
244 =	Eingabe/Ausgabe;
300 =	Aufgenommenes Bild;
350 =	Bild;
302 =	Zentrum;
308, 312 =	Person;
310 =	Gesicht;
322, 324 =	Auge;
330 =	Mund;
404A, 404B =	Grenze;
406 =	Richtung;
408A, 408B =	Mikrofon;
410A, 410B =	Abstand;
502 =	Bestimmen eines Abstands eines Gesichts;
504 =	Bestimmen einer Größe einer Anzeige;
506 =	Skalieren des Bilds;

Schutzansprüche

1. Computerprogramm, das Anweisungen umfasst, die dann, wenn sie auf wenigstens einem Prozessor ausgeführt werden, veranlassen, dass ein Computersystem Schritte ausführt, die folgendes umfassen:
Bestimmen eines Abstands eines Gesichts einer Person von einer Kamera, die ein Bild des Gesichts der Person aufnahm;
Bestimmen einer Größe einer Anzeige in Kommunikation mit der Kamera; und

Skalieren des Bilds basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige.

2. Computerprogramm nach Anspruch 1, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:

Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person;
Messen einer Breite zwischen dem linken Auge und dem rechten Auge;
Teilen einer vorbestimmten Breite durch eine gemessene Breite, um ein Breitenverhältnis zu bestimmen; und
Multiplizieren des Breitenverhältnisses mit einem vorbestimmten Abstand.

3. Computerprogramm nach Anspruch 1, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:

Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person;
Bestimmen eines Mittelpunkts zwischen den Augen von dem linken Auge und dem rechten Auge;
Erkennen eines Mundes im Bild des Gesichts der Person;
Bestimmen eines Mundmittelpunkts des Mundes;
Messen einer Höhe vom Mundmittelpunkt zum Mittelpunkt zwischen Augen;
Teilen einer vorbestimmten Höhe durch die gemessene Höhe, um ein Höhenverhältnis zu bestimmen; und
Multiplizieren des Höhenverhältnisses mit einem vorbestimmten Abstand.

4. Computerprogramm nach Anspruch 1, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:

Messen einer Breite zwischen einem linken Auge und einem rechten Auge im Bild des Gesichts der Person;
Messen einer Höhe zwischen einem Mittelpunkt zwischen Augen und einem Mundmittelpunkt im Bild des Gesichts der Person;
Teilen einer größeren der Breite und der Höhe durch eine vorbestimmte Breite, wenn die Breite größer ist, und durch eine vorbestimmte Höhe, wenn die Höhe größer ist, um ein Verhältnis zu erzeugen; und
Multiplizieren des Verhältnisses mit einem vorbestimmten Abstand.

5. Computerprogramm nach Anspruch 1, wobei das Bestimmen der Größe der Anzeige ein Bestimmen der Größe der Anzeige basierend auf einem Anzeigengrößenindikator enthält, der in einer von einer mit der Anzeige assoziierten Computervorrichtung empfangenen Hypertext-Übertragungsprotokoll-(HTTP-)Nachricht enthalten ist.

6. Computerprogramm nach Anspruch 1, wobei das Skalieren des Bilds ein Skalieren des Bilds zu ei-

ner Größe auf der Anzeige enthält, die zu einer Größe passt, dass das Gesicht der Person einem Betrachter bei einem Standort der Anzeige aus dem bestimmten Abstand des Gesichts der Person von der Kamera erscheinen würde.

7. Computerprogramm nach Anspruch 1, das weiterhin ein Erkennen des Gesichts der Person als einen aktiven Sprecher basierend auf einer Zeitdifferenz zwischen Audioeingaben umfasst, die anzeigen, dass ein Standort des aktiven Sprechers mit einem bestimmten Standort des Gesichts der Person übereinstimmt.

8. Computerprogramm nach Anspruch 7, wobei der Standort des Gesichts der Person basierend auf dem bestimmten Abstand des Gesichts der Person und einer Richtung des Gesichts der Person von der Kamera bestimmt wird.

9. Computerprogramm nach Anspruch 1, das weiterhin ein Erkennen des Gesichts der Person als einen aktiven Sprecher basierend auf einem Erkennen einer Lippenbewegung des Gesichts der Person umfasst.

10. Computerprogramm nach Anspruch 1, das weiterhin ein Erkennen des Gesichts der Person als einen aktiven Sprecher basierend auf einem Erkennen von Blicken der Augen von anderen Personen in Richtung zu dem Gesicht der Person umfasst.

11. Computerprogramm nach Anspruch 1, das weiterhin ein Senden des skalierten Bilds zur Anzeige umfasst.

12. Computerprogramm nach Anspruch 1, das weiterhin ein Empfangen des aufgenommenen Bilds von der Kamera umfasst.

13. Computerprogramm nach Anspruch 1, wobei das Verfahren durch ein Computersystem durchgeführt wird, das die Anzeige steuert.

14. Computerprogramm nach Anspruch 1, wobei das Verfahren durch ein Computersystem durchgeführt wird, das die Kamera steuert.

15. Nichtflüchtiges computerlesbares Speichermedium, das darauf gespeicherte Anweisungen umfasst, die dann, wenn sie durch wenigstens einen Prozessor ausgeführt werden, konfiguriert sind, um zu veranlassen, dass ein Computersystem wenigstens:

einen Abstand eines Gesichts einer Person von einer Kamera, die ein Bild des Gesichts der Person aufnahm, bestimmt;
eine Größe einer Anzeige in Kommunikation mit der Kamera bestimmt; und

das Bild basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige skaliert.

16. Nichtflüchtiges computerlesbares Speichermedium nach Anspruch 15, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält: Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person; Messen einer Breite zwischen dem linken Auge und dem rechten Auge; Teilen einer vorbestimmten Breite durch eine gemessene Breite, um ein Breitenverhältnis zu bestimmen; und Multiplizieren des Breitenverhältnisses mit einem vorbestimmten Abstand.

17. Nichtflüchtiges computerlesbares Speichermedium nach Anspruch 15, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält: Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person; Bestimmen eines Mittelpunkts zwischen den Augen von dem linken Auge und dem rechten Auge; Erkennen eines Mundes im Bild des Gesichts der Person; Bestimmen eines Mundmittelpunkts des Mundes; Messen einer Höhe vom Mundmittelpunkt zum Mittelpunkt zwischen Augen; Teilen einer vorbestimmten Höhe durch die gemessene Höhe, um ein Höhenverhältnis zu bestimmen; und Multiplizieren des Höhenverhältnisses mit einem vorbestimmten Abstand.

18. Nichtflüchtiges computerlesbares Speichermedium nach Anspruch 15, wobei das nichtflüchtige computerlesbare Speichermedium auf einem Computersystem gespeichert ist, das die Anzeige steuert.

19. Nichtflüchtiges computerlesbares Speichermedium nach Anspruch 15, wobei das nichtflüchtige computerlesbare Speichermedium auf einem Computersystem gespeichert ist, das die Kamera steuert.

20. System, umfassend:
eine Kamera, die konfiguriert ist, um Bilder aufzunehmen;
eine Anzeige, die konfiguriert ist, um Bilder zu präsentieren;
wenigstens einen Prozessor, der konfiguriert ist, um Anweisungen auszuführen; und
ein nichtflüchtiges computerlesbares Speichermedium, das darauf gespeicherte Anweisungen umfasst, die dann, wenn sie durch den wenigstens einen Prozessor ausgeführt werden, konfiguriert sind, um zu veranlassen, dass das System wenigstens:
einen Abstand eines Gesichts einer Person von einer Kamera bestimmt, die ein Bild des Gesichts der Person aufnahm;

eine Größe einer Anzeige bestimmt, wobei die Anzeige in Kommunikation mit der Kamera ist; und
das Bild basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige skaliert.

21. Computerprogramm zum Ermöglichen einer Videokonferenz, das Anweisungen umfasst, die dann, wenn sie auf wenigstens einem Prozessor ausgeführt werden, veranlassen, dass ein Computersystem Schritte ausführt, die folgendes umfassen:
Bestimmen eines Abstands eines Gesichts einer Person von einer Kamera, die ein Bild des Gesichts der Person aufnahm;
Bestimmen einer Größe einer Anzeige in Kommunikation mit der Kamera; und
Skalieren des Bilds basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige.

22. Computerprogramm nach Anspruch 21, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:
Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person;
Messen einer Breite zwischen dem linken Auge und dem rechten Auge;
Teilen einer vorbestimmten Breite durch eine gemessene Breite, um ein Breitenverhältnis zu bestimmen; und
Multiplizieren des Breitenverhältnisses mit einem vorbestimmten Abstand.

23. Computerprogramm nach Anspruch 21, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:
Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person;
Bestimmen eines Mittelpunkts zwischen den Augen von dem linken Auge und dem rechten Auge;
Erkennen eines Mundes im Bild des Gesichts der Person;
Bestimmen eines Mundmittelpunkts des Mundes;
Messen einer Höhe vom Mundmittelpunkt zum Mittelpunkt zwischen Augen;
Teilen einer vorbestimmten Höhe durch die gemessene Höhe, um ein Höhenverhältnis zu bestimmen; und
Multiplizieren des Höhenverhältnisses mit einem vorbestimmten Abstand.

24. Nichtflüchtiges computerlesbares Speichermedium, das darauf gespeicherte Anweisungen umfasst, zum Ermöglichen einer Videokonferenz, die dann, wenn sie durch wenigstens einen Prozessor ausgeführt werden, konfiguriert sind, um zu veranlassen, dass ein Computersystem wenigstens:
einen Abstand eines Gesichts einer Person von einer Kamera, die ein Bild des Gesichts der Person aufnahm, bestimmt;

eine Größe einer Anzeige in Kommunikation mit der Kamera bestimmt; und
das Bild basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige skaliert.

ge gemessen wird, umgekehrt proportional zu der bestimmten Größe der Anzeige ist.

Es folgen 6 Seiten Zeichnungen

25. Nichtflüchtiges computerlesbares Speichermedium nach Anspruch 24, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:
Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person;
Messen einer Breite zwischen dem linken Auge und dem rechten Auge;
Teilen einer vorbestimmten Breite durch eine gemessene Breite, um ein Breitenverhältnis zu bestimmen; und
Multiplizieren des Breitenverhältnisses mit einem vorbestimmten Abstand.

26. Nichtflüchtiges computerlesbares Speichermedium nach Anspruch 24, wobei das Bestimmen des Abstands des Gesichts der Person folgendes enthält:
Erkennen eines linken Auges und eines rechten Auges im Bild des Gesichts der Person;
Bestimmen eines Mittelpunkts zwischen den Augen von dem linken Auge und dem rechten Auge;
Erkennen eines Mundes im Bild des Gesichts der Person;
Bestimmen eines Mundmittelpunkts des Mundes;
Messen einer Höhe vom Mundmittelpunkt zum Mittelpunkt zwischen Augen;
Teilen einer vorbestimmten Höhe durch die gemessene Höhe, um ein Höhenverhältnis zu bestimmen; und
Multiplizieren des Höhenverhältnisses mit einem vorbestimmten Abstand.

27. System zum Ermöglichen einer Videokonferenz, wobei das System folgendes umfasst:
eine Kamera, die konfiguriert ist, um Bilder aufzunehmen;
eine Anzeige, die konfiguriert ist, um Bilder zu präsentieren;
wenigstens einen Prozessor, der konfiguriert ist, um Anweisungen auszuführen; und
ein nichtflüchtiges computerlesbares Speichermedium, das darauf gespeicherte Anweisungen umfasst, die dann, wenn sie durch den wenigstens einen Prozessor ausgeführt werden, konfiguriert sind, um zu veranlassen, dass das System wenigstens:
einen Abstand eines Gesichts einer Person von einer Kamera bestimmt, die ein Bild des Gesichts der Person aufnahm;
eine Größe der Anzeige bestimmt, wobei die Anzeige in Kommunikation mit der Kamera ist; und
das Bild basierend auf dem bestimmten Abstand des Gesichts der Person und der bestimmten Größe der Anzeige skaliert, wobei das skalierte Bild eine Größe hat, die dann, wenn sie als ein Anteil der Anzei-

Anhängende Zeichnungen

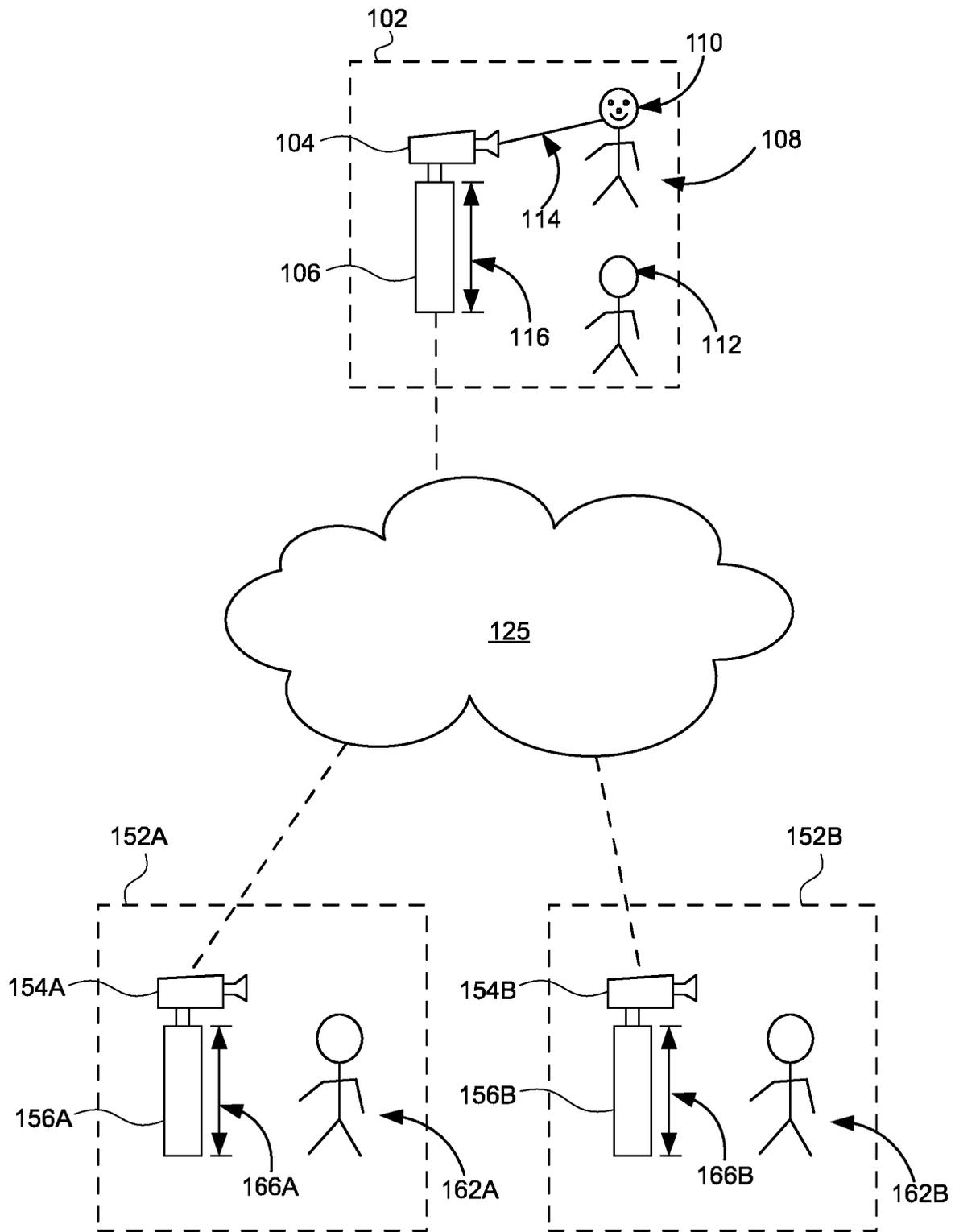


FIG. 1

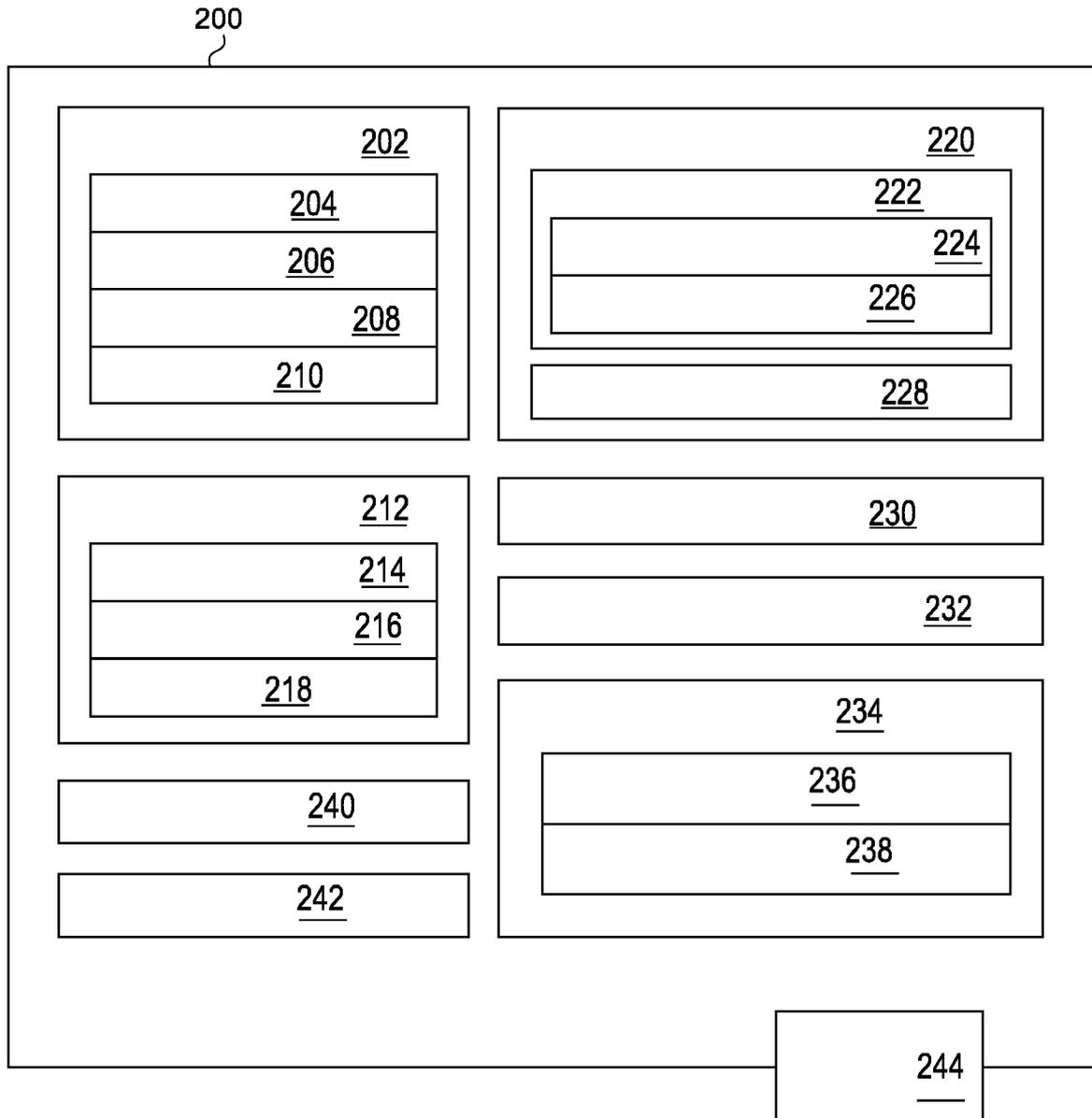


FIG. 2

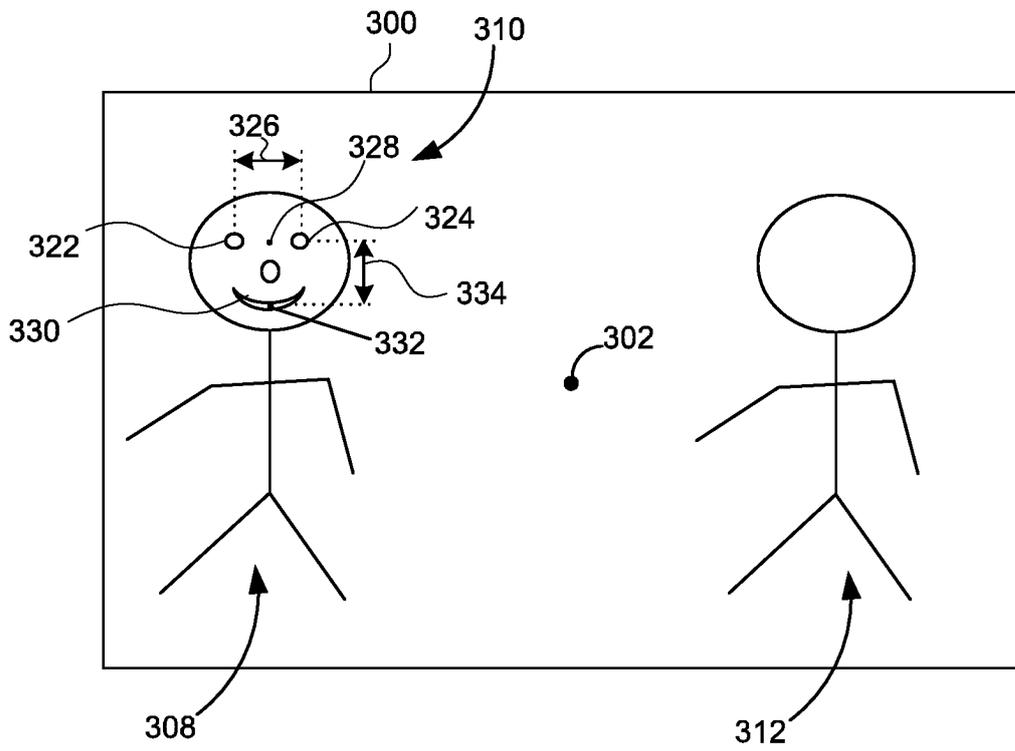


FIG. 3A

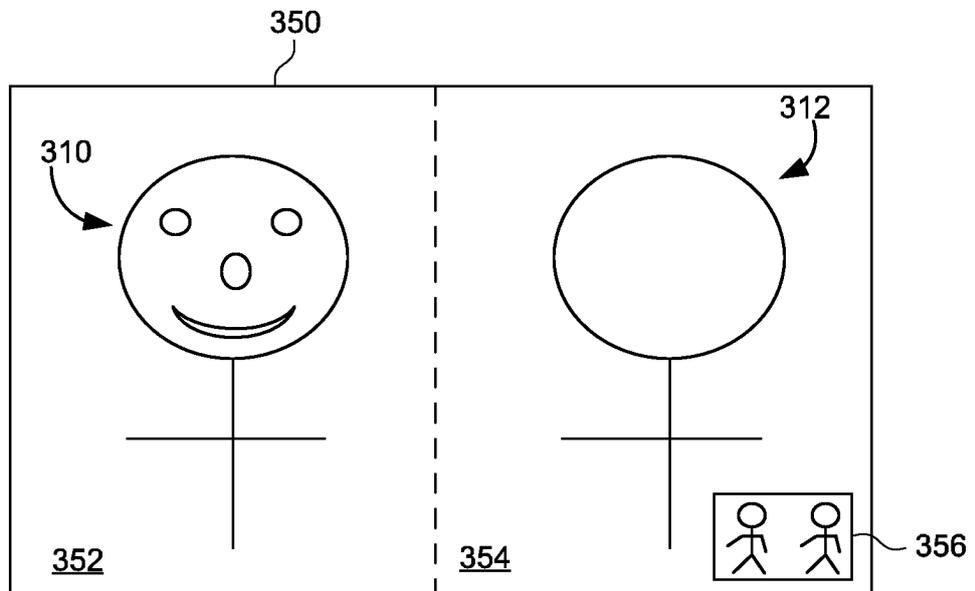


FIG. 3B

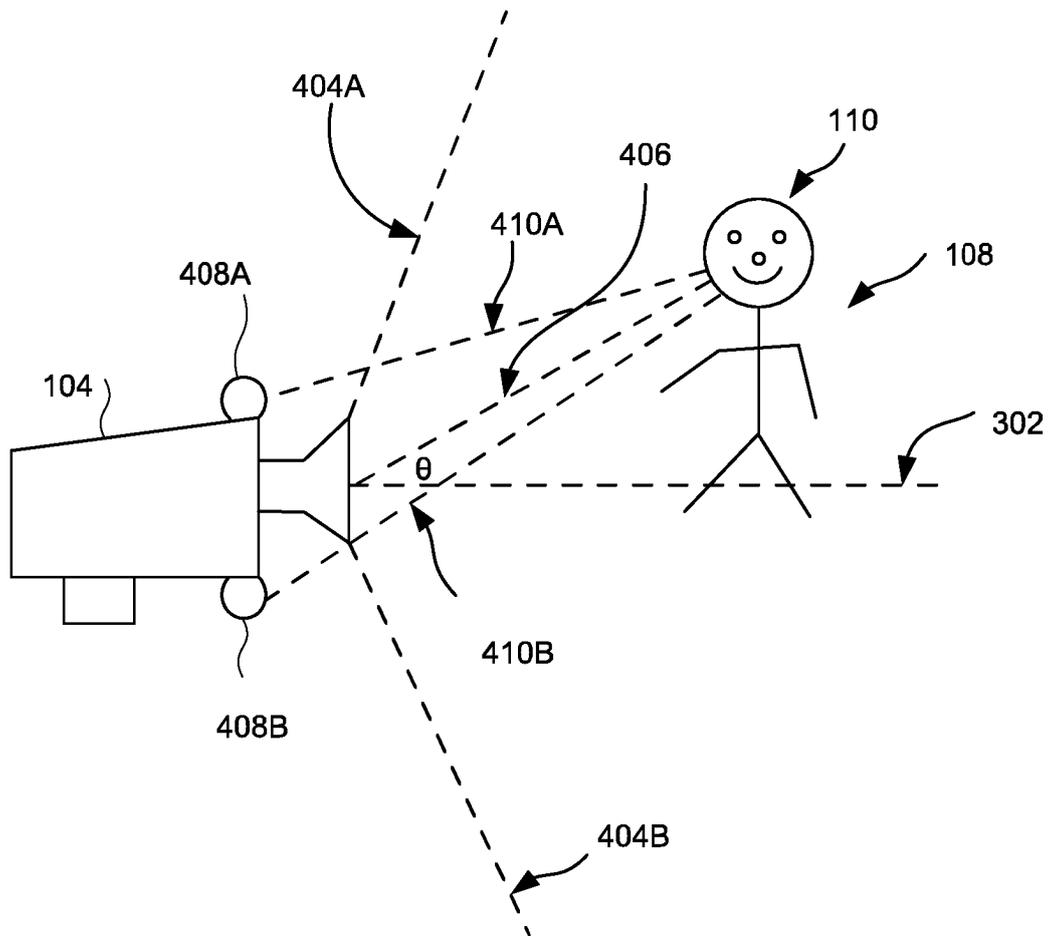


FIG. 4

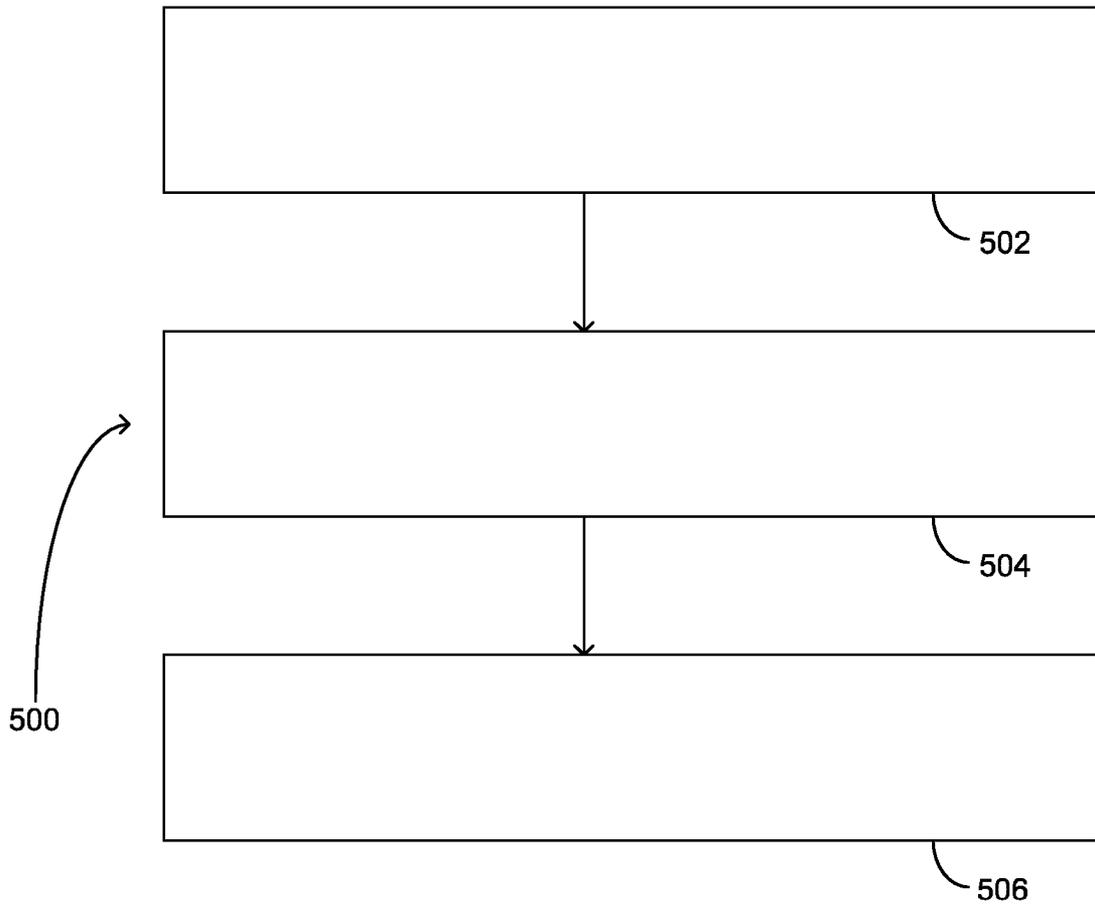


FIG. 5

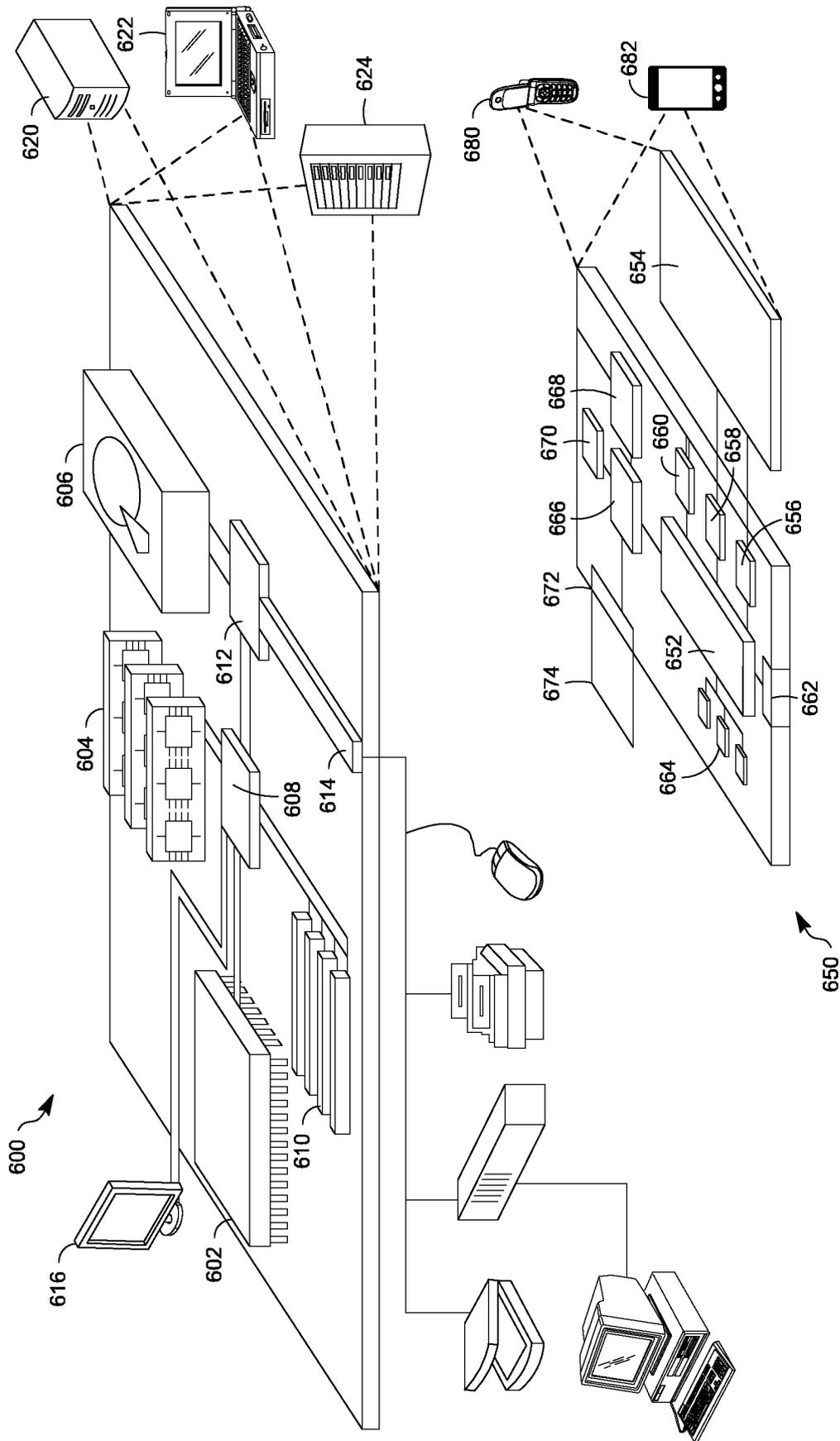


FIG. 6